Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*
Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University
URIE BRONFENBRENNER
Cornell University
ROGER BROWN
Massachusetts Institute of Technology
DONALD T. CAMPBELL
Northwestern University
LAUNOR F. CARTER
System Development Corporation
RICHARD CHRISTIE
Columbia University
MORTON DEUTSCH
Bell Telephone Laboratories
I. E. FARBER
State University of Iowa
DONALD W. FISKE
University of Chicago
N. L. GAGE
University of Illinois
ROGER W. HEYNS
University of Michigan
ROBERT R. HOLT
New York University
IRVING L. JANIS
Yale University
HAROLD H. KELLEY
University of Minnesota
LEONARD S. KOGAN
Community Service Society of New York
RICHARD S. LAZARUS
University of California, Berkeley
DANIEL J. LEVINSON
Harvard Medical School
DAVID C. McCLELLAND
Harvard University
LEO POSTMAN
University of California, Berkeley
ELIOT H. RODNICK
Duke University
NEVITT SANFORD
University of California, Berkeley
SEYMOUR SARASON
Yale University
RICHARD L. SOLOMON
University of Pennsylvania

Managing Editor: ARTHUR C. HOFFMAN
Promotion Manager: HELEN ORR

VOLUME 61

1960

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1961 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 61

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

PAGE
PORT, GORDON W. **The Open System in Personality Theory** 301
ERT, RICHARD, & HABER, RALPH NORMAN. **Anxiety in Academic Achievement Situations** 207
ROCCHI, JOHN, PARSONS, OSCAR A., & DICKOFF, HILDA. **Changes in Self-Ideal Discrepancy in Repressors and Sensitizers** 67
EY, LEO B. **The Indirect Representation of Sexual Stimuli by Schizophrenic and Normal Subjects** 424
ITAGE, STEWART G. See KING, GERALD F.
LDWIN, ALFRED L. See LEVIN, HARRY.
YLEY, NANCY. See SCHAEFER, EARL S.
CKER, WESLEY C. **Cortical Inhibition and Extraversion-Introversion** 52
NJAMIN, LAWRENCE. See KATZ, IRWIN.
OKBINDER, LAWRENCE J. See RECHTSCHAFFEN, ALLAN.
BREKSTAD, ARNE. See HAGGARD, ERNEST A.
BUTMAN, JEAN. See SCHLESINGER, LAWRENCE.
CARLSON, EARL R., & CARLSON, RAE. **Male and Female Subjects in Personality Research** 482
CARLSON, RAE. See CARLSON, EARL R.
CHOWN, SHEILA M. **A Factor Analysis of the Wesley Rigidity Inventory: Its Relationship to Age and Nonverbal Intelligence** 491
COMMINS, WILLIAM D. See REILLY, MARY ST. ANNE.
COTTRELL, NICKOLAS B. See STOTLAND, EZRA.
DAS, J. P. **Prestige Effects in Body-Sway Suggestibility** 487
DAS, J. P. See KANUNGO, R.
DAVIDS, ANTHONY, & SILVERMAN, MORTON. **A Psychological Case Study of Death during Pregnancy** 287
DAVIS, KEITH E., & JONES, EDWARD E. **Changes in Interpersonal Perception as a Means of Reducing Cognitive Dissonance** 402
DEUTSCH, MORTON. **Trust, Trustworthiness, and the F Scale** 138
DEUTSCH, MORTON, & KRAUSS, ROBERT M. **The Effect of Threat upon Interpersonal Bargaining** 181
DICKOFF, HILDA. See ALTROCCHI, JOHN.
DRECHSLER, ROBERT J. **Affect-Stimulating Effects of Colors** 323
DUGAN, ROBERT D. See LEVY, LEON H.
ENDLER, NORMAN S. **Social Conformity in Perception of the Autokinetic Effect** 489
ERIKSEN, CHARLES W. See FUHRER, MARCUS J.
ERVIN, SUSAN M., & FOSTER, GARRETT. **The Development of Meaning in Children's Descriptive Terms** 271
FARINA, AMERIGO. **Patterns of Role Dominance and Conflict in Parents of Schizophrenic Patients** 31
FIELD, WILLIAM L. See STEINER, IVAN D.
FOSHEE, JAMES G. See SIEGEL, PAUL S.
FOSTER, GARRETT. See ERVIN, SUSAN M.
FRANC, DOUGLAS E. See ROSENBAUM, MILTON E.
FRENCH, JOHN R. P., JR., MORRISON, H. WILLIAM, & LEVINGER, GEORGE. **Coercive Power and Forces Affecting Conformity** 93
FUHRER, MARCUS J., & ERIKSEN, CHARLES W. **The Unconscious Perception of the Meaning of Verbal Stimuli** 432
GALLWEY, MARY. See LEVIN, HARRY.
GARDNER, RILEY W., & LONG, ROBERT I. **The Stability of Cognitive Controls** 485
GARDNER, RILEY W. See HOLZMAN, PHILIP S.
GETZELS, J. W., & JACKSON, P. W. **Occupational Choice and Cognitive Functioning: Career Aspirations of Highly Intelligent and of Highly Creative Adolescents** 119
GORDON, JESSE E. See PATEL, AMBALAL S.
GORHAM, DONALD R. See MORAN, LOUIS J.
HABER, RALPH NORMAN. See ALPERT, RICHARD.
HAGGARD, ERNEST A., BREKSTAD, ARNE, & SKARD, ÅSE GRUDA. **On the Reliability of the Anamnestic Interview** 311
HOLLANDER, E. P. **Competence and Conformity in the Acceptance of Influence** 365

PAGE

HOLT, ROBERT R. See PINE, FRED.
HOLTZMAN, WAYNE W. See MORAN, LOUIS J.
HOLZMAN, PHILIP S., & GARDNER, RILEY W. Leveling-Sharpening and Memory Organization 176
IZARD, CARROLL E. Personality Similarity and Friendship 47
IZARD, CARROLL E. Personality Similarity, Positive Affect, and Interpersonal Attraction 484
JACKSON, JAY M. See SCHLESINGER, LAWRENCE.
JACKSON, P. W. See GETZELS, J. W.
JONES, EDWARD E. See DAVIS, KEITH E.
JONES, EDWARD E. See STRICKLAND, LLOYD H.
KAGAN, JEROME, & LEMKIN, JUDITH. The Child's Differential Perception of Parental Attributes 440
KANFER, FREDERICK H. Verbal Rate, Eyeblink, and Content in Structured Psychiatric Interviews .. 341
KANUNGO, R., & DAS, J. P. Differential Learning and Forgetting as a Function of the Social Frame of Reference 82
KARON, BERTRAM P. A Clinical Note on the Specific Nature of an "Oral" Trauma 480
KATZ, IRWIN, & BENJAMIN, LAWRENCE. Effects of White Authoritarianism in Biracial Work Groups .. 448
KING, GERALD F., ARMITAGE, STEWART G., & TILTON, JOHN R. A Therapeutic Approach to Schizophrenics of Extreme Pathology: An Operant-Interpersonal Method 276
KRAUSS, ROBERT M. See DEUTSCH, MORTON.
KUETHE, JAMES L. Acquiescent Response Set and the Psychasthenia Scale: An Analysis via the *Aussage* Experiment 319
LAFFAL, JULIUS. The Contextual Associates of Sun and God in Schreber's Autobiography 474
LAING, GORDON. See STOTLAND, EZRA.
LEMKIN, JUDITH. See KAGAN, JEROME.
LEVIN, HARRY, BALDWIN, ALFRED L., GALLWEY, MARY, & PAIVIO, ALLAN. Audience Stress, Personality, and Speech 469
LEVINGER, GEORGE. See FRENCH, JOHN R. P., JR.
LEVY, LEON H. Context Effects in Social Perception 295
LEVY, LEON H., & DUGAN, ROBERT D. A Constant Error Approach to the Study of Dimensions of Social Perception 21
LINDZEY, GARDNER, LYKKEN, DAVID T., & WINSTON, HARVEY D. Infantile Trauma, Genetic Factors, and Adult Temperament 7
LONG, ROBERT I. See GARDNER, RILEY W.
LOTT, ALBERT J. See LOTT, BERNICE EISMAN.
LOTT, BERNICE EISMAN, & LOTT, ALBERT J. The Formation of Positive Attitudes toward Group Members 297
LYKKEN, DAVID T. See LINDZEY, GARDNER.
MANGAN, G. L., QUARTERMAIN, D., & VAUGHAN, G. M. Taylor MAS and Group Conformity Pressure .. 14
MATARAZZO, JOSEPH D., SASLOW, GEORGE, & PAREIS, E. NELSON. Verbal Conditioning of Two Response Classes: Some Methodological Considerations 19
MORAN, LOUIS J., GORHAM, DONALD R., & HOLTZMAN, WAYNE W. Vocabulary Knowledge and Usage of Schizophrenic Subjects: A Six-Year Follow-Up 24
MORRISON, H. WILLIAM. See FRENCH, JOHN R. P., JR.
MORRISSETTE, JULIAN. See ZAJONC, ROBERT B.
O'CONNELL, WALTER E. The Adaptive Functions of Wit and Humor 26
PAIVIO, ALLAN. See LEVIN, HARRY.
PAREIS, E. NELSON. See MATARAZZO, JOSEPH D.
PARSONS, OSCAR A. See ALTROCCHI, JOHN.
PATEL, AMBALAL S., & GORDON, JESSE E. Some Personal and Situational Determinants of Yielding to Influence 41
PEAK, HELEN. The Effect of Aroused Motivation on Attitudes 46
PERVIN, LAWRENCE A. Rigidity in Neurosis and General Personality Functioning 38
PHILLIPS, LESLIE. See ZIGLER, EDWARD.
PINE, FRED, & HOLT, ROBERT R. Creativity and Primary Process: A Study of Adaptive Regression ... 37
PROTHRO, E. TERRY. Patterns of Permissiveness among Preliterate Peoples 1
QUARTERMAIN, D. See MANGAN, G. L.
RAO, KUNAPULI U., & RUSSELL, ROGER W. Effects of Stress on Goal Setting Behavior 3
RECHTSCHAFFEN, ALLAN, & BOOKBINDER, LAWRENCE J. Introversion-Extraversion and Kinesthetic Aftereffects 4

PAGE

Reilly, Mary St. Anne, Commins, William D., & Stefic, Edward C. **The Complementarity of Personality Needs in Friendship Choice** 292
Roen, Sheldon R. **Personality and Negro-White Intelligence** 148
Rosenbaum, Milton E., & Franc, Douglas E. **Opinion Change as a Function of External Commitment and Amount of Discrepancy from the Opinion of Another** 15
Rothstein, Ralph. **Authoritarianism and Men's Reactions to Sexuality and Affection in Women** .. 329
Russell, Roger W. See Rao, Kunapuli U.
Saslow, George. See Matarazzo, Joseph D.
Saumer, Ellen. See Secord, Paul F.
Schaefer, Earl S., & Bayley, Nancy. **Consistency of Maternal Behavior from Infancy to Preadolescence** 1
Schiller, Peter H. See Wiener, Morton.
Schlesinger, Lawrence, Jackson, Jay M., & Butman, Jean. **Leader-Member Interaction in Management Committees** 360
Schooler, Carmi, & Spohn, Herbert E. **The Susceptibility of Chronic Schizophrenics to Social Influence in the Formation of Perceptual Judgments** 348
Secord, Paul F., & Saumer, Ellen. **Identifying Jewish Names: Does Prejudice Increase Accuracy?** . 144
Showel, Morris. **Interpersonal Knowledge and Rated Leader Potential** 87
Siegel, Paul S., & Foshee, James G. **Molar Variability in the Mentally Defective** 141
Silverman, Morton. See Davids, Anthony.
Skard, Åse Gruda. See Haggard, Ernest A.
Smith, Anthony J. **The Attribution of Similarity: The Influence of Success and Failure** 419
Smith, William P. See Strickland, Lloyd H.
Solomon, Leonard. **The Influence of Some Types of Power Relationships and Game Strategies upon the Development of Interpersonal Trust** 223
Solomon, Leonard. See Wilensky, Harold.
Spohn, Herbert E. See Schooler, Carmi.
Stefic, Edward C. See Reilly, Mary St. Anne.
Steiner, Ivan D., & Field, William L. **Role Assignment and Interpersonal Influence** 239
Stotland, Ezra, Cottrell, Nickolas B., & Laing, Gordon. **Group Interaction and Perceived Similarity of Members** 335
Strickland, Lloyd H., Jones, Edward E., & Smith, William P. **Effects of Group Support on the Evaluation of an Antagonist** 73
Suci, George J. **A Comparison of Semantic Structures in American Southwest Culture Groups** ... 25
Talland, George A. **A Microgenetic Approach to the Amnesic Syndrome** 255
Tilton, John R. See King, Gerald F.
Triandis, Harry C., & Triandis, Leigh Minturn. **Race, Social Class, Religion, and Nationality as Determinants of Social Distance** 110
Triandis, Leigh Minturn. See Triandis, Harry C.
Underwood, Harold Wayne. **The Validity of Hypnotically Induced Visual Hallucinations** 39
Vaughan, G. M. See Mangan, G. L.
Wagoner, Robert A. **Differences in Response Latency and Response Variability between High and Low Anxiety Subjects in a Flicker-Fusion Task** 355
Wallach, Michael A. **Two Correlates of Symbolic Sexual Arousal: Level of Anxiety and Liking for Esthetic Material** 396
Wiener, Morton, & Schiller, Peter H. **Subliminal Perception or Perception of Partial Cues** 124
Wilensky, Harold, & Solomon, Leonard. **Characteristics of Untestable Chronic Schizophrenics** ... 155
Winston, Harvey D. See Lindzey, Gardner.
Wrightsman, Lawrence S., Jr. **Effects of Waiting with Others on Changes in Level of Felt Anxiety** . 216
Zahn, Theodore P. **Size Estimation of Pictures Associated with Success and Failure as a Function of Manifest Anxiety** 457
Zajonc, Robert B. **The Process of Cognitive Tuning in Communication** 159
Zajonc, Robert B., & Morrissette, Julian. **The Role of Uncertainty in Cognitive Change** 168
Zigler, Edward, & Phillips, Leslie. **Social Effectiveness and Symptomatic Behaviors** 231
Zipf, Sheila G. **Resistance and Conformity under Reward and Punishment** 102

Volume 61 July 1960 Number 1

Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*
Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

DONALD W. FISKE
University of Chicago

N. L. GAGE
University of Illinois

ROGER W. HEYNS
University of Michigan

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

RICHARD S. LAZARUS
University of California, Berkeley

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California, Berkeley

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California, Berkeley

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
University of Pennsylvania

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 900 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

M. BREWSTER SMITH
Department of Psychology
University of California
Berkeley 4, California

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1960 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 1-6

CONSISTENCY OF MATERNAL BEHAVIOR FROM INFANCY TO PREADOLESCENCE

EARL S. SCHAEFER AND NANCY BAYLEY
National Institute of Mental Health, Bethesda, Maryland

CURRENT social psychological theories of personality development are concerned with the cumulative effect upon the child of the social and emotional behaviors of the parent throughout the period of childhood. Interest in these cumulative effects leads to the question whether from observations or interviews on parental behavior at one developmental stage it is possible to determine consistent, typical parental behavior that existed earlier and will continue throughout childhood. This problem should be of special concern to retrospective studies in which parents of contrasting groups are either observed or interviewed and inferences are then made that current differences in parental behavior existed prior to the children's behavior, or that they contributed to its development.

A second problem in research concerned with evaluating the influence of maternal behavior upon personality development is that of the validity of both observation of mother-child interaction and of the interview as sources of data on the mother's typical behavior with the child. Observation of mother-child interaction is more feasible during infancy when the mother usually maintains close contact with the child and interacts relatively spontaneously with the infant in the presence of the observer. At later ages, when the child's contact with the mother is usually reduced and when the presence of the observer often limits the spontaneity of mother-child interaction, observation is often replaced by an interview with the mother from which inferences are made about her behavior with the child. Little is known about the relationship between observation and interview as sources of data on maternal behavior.

Available longitudinal data permitted tests of two hypotheses that are suggested by these problems:

1. A mother's behavior with her infant is significantly related to her behavior with the same child at preadolescence.

2. Data on a mother's behavior with her child obtained from an interview are significantly related to data obtained from direct observation of mother-child interaction.

The available data led to a research design that unfortunately confounded these two hypotheses. Only two sets of data were available, data quantified from records of observations during the period of one month to three years of the child's life (infancy) and data quantified from interviews with the mother at child's age of from 9 to 14 years (preadolescence).[1] If these two sets of data are significantly related, both hypotheses must be accepted. If, however, the two sets of data are not significantly related, the inconsistency of maternal behavior might be attributed either to true changes in behavior across this time span, to the lack of validity of either the observational or interview data, or both.

METHOD

Subjects and Data

A description of the subjects (*Ss*) and of the research in which these data were collected has been reported elsewhere (Jones & Bayley, 1941). Both infancy observation data and preadolescent interview data were available for mothers of 15 boys and 16 girls. Because of the small size of the two groups the data on mothers of boys and of girls were combined for the analysis of individual scales. The consistency of the two major dimensions were computed both for the combined groups and separately for the two sexes.

The observations were collected by an observer[2] who was present during the testing sessions with mother and child in the University of California Institute of Child Welfare during the first three years of the child's life. Tests with the children were scheduled every month for the first 15 months and every 3 months thereafter through the age of 3 years. The later interviews were made by a skilled psychologist[3] who had no knowledge of the earlier observations. The mothers were interviewed during either one or two visits in the home between the child's ages of nine to fourteen years. The interview was focused upon current family behavior rather than upon earlier events. A comprehensive clinical report of the interviews was written after the home visit. The early observations and later interviews were employed (Schaefer, Bell, & Bayley, 1959) in developing measures for a conceptual scheme that emphasized molar social psychological variables (see Table 1). The conceptual scheme was first used to quantify the birth to 3-year observations and then was modified slightly to make the variables appropriate for the 9- to 14-year interview data. Each of the two sets of data was rated by 3 judges, although only one of these judges rated both sets.

---

[1] The age of the children at the time of the preadolescent interview ranged from 9 to 14 years, with a median and mode of 13 years.

[2] Most of the observations were made by a graduate student in psychology, H. Elizabeth Neall. Some were made by one or other of the examiners.

[3] Pearl Bretnall Meissner.

The median combined interjudge reliability (product-moment correlations) for the ratings of the observation data was .82 and the median reliability for the ratings of the interview data was .85. The consistency correlations for the one judge who rated both sets of data were not significantly different from the consistency correlations between independent raters. Only the consistencies of the combined ratings of the three judges will be reported.

In an effort to organize the variables that have been described, the variables that were directly related to the social and emotional behavior of the mother toward her child were intercorrelated for each set of data. Additional variables that were rated but were not included in this analysis are descriptive of the mother's personality, or are related to situational circumstances. An organization of the maternal variables using the circumplex model (Guttman, 1954) and factor analysis (Thurstone, 1947), revealed an ordering within a two-dimensional space (Schaefer, 1959). The two dimensions were labeled Love versus Hostility and Autonomy versus Control for both sets of data.

RESULTS

The infancy observation variables were correlated with the preadolescent interview variables using Pearson produce-moment correlations. Both sets of variables are reported in Table 1 in the circumplex order that was found in the previous analysis (Schaefer, 1959). The combined interrater reliabilities are also reported. All correlations above .36 are significant at the .05 level and all correlations above .47 are significant at the .01 level using a two-tailed test for this sample of 31 *S*s. The significant correlations tend to occur between variables that define a dimension of Love versus Hostility while the insignificant correlations are between variables that define a dimension of Autonomy versus Control at each age level.

As a method of summarizing this finding, the variables that had high intercorrelations and high loadings on a dimension were organized into cluster scores. The Autonomy versus Control dimension was defined for the period of infancy by autonomy, excessive contact, intrusiveness, and fostering dependency and for preadolescence by autonomy, excessive contact and intrusiveness. The Love versus Hostility dimension was defined for the early observations by cooperativeness, positive evaluation of the child, equalitarianism, expression of affection, ignoring, irritability, punishment, punitiveness, perceives the child as a burden, use of fear to control the child, and strictness, and for the later interviews by cooperativeness, positive evaluation of the child, equalitarianism, expression of affection, positive mother-child relationship, ignoring, irritability, punitiveness, use of fear to control the child and rejection of the homemaking role. Sums of the unweighted scores for the variables in a cluster were determined for each of the 31 *S*s at both age levels and these two levels were then intercorrelated, using Pearson product-moment correlations. The intercorrelations and estimated combined interrater reliabilities for each of the clusters are reported in Table 2. The correlation for the combined sexes between the Love versus Hostility clusters is .68, significant beyond the .01 level, while the correlation between the Autonomy versus Control clusters is .26, not significantly different from zero. The results for the two sexes do not differ significantly. The correlation for the Love–Hostility axis for boys is .73 and for girls is .61 and both are statistically significant. The consistency correlation for the Autonomy–Control axis for boys is .18 and for girls is .35 and neither one is statistically significant. These results summarize and confirm the findings reported in Table 1. The correlation of .34 between the Love-Hostility and Autonomy-Control axes at preadolescence merely indicates that the method of clustering variables did not give independent dimensions.

Although Tables 1 and 2 adequately summarize the data on consistency of maternal behavior for variables that clearly relate to the behavior of the mother with the individual child, other variables of the infancy data were analyzed to determine the extent to which they are related to the preadolescent ratings of positiveness of mother-child interaction. The significant correlations of the early ratings

TABLE 1

CORRELATION OF CIRCUMPLEX-ORDERED INFANCY OBSERVATION RATINGS OF MATERNAL BEHAVIOR WITH CIRCUMPLEX-ORDERED PREADOLESCENT INTERVIEW RATINGS OF MATERNAL BEHAVIOR

Preadolescent Interview Ratings	Reliability	Infancy Observation Ratings																
		Autonomy	Positive Evaluation	Equalitarianism	Expression of Affection	Emotional Involvement	Fostering Dependency	Excessive Contact	Achievement Demand	Concern About Health	Intrusiveness	Anxiety of Mother	Irritability	Punishment	Use of Fear to Control	Strictness	Perceives as Burden	Punitiveness	Ignoring
		80	82	84	83	87	85	67	81	81	76	83	82	87	86	78	88	78	89
Autonomy	77	29	01	13	−08	−04	−09	−26	−19	−25	−23	−01	06	−07	−11	−14	08	01	15
Positive Mother-Child Relationship	88	05	*48**	*58*	*39*	25	16	20	03	10	−07	−10	*−51*	*−42*	*−52*	*−36*	*−53*	*−54*	*−41*
Equalitarianism	83	09	*52*	*56*	*45*	28	16	19	09	08	−12	−08	*−53*	*−41*	*−51*	*−39*	*−57*	*−60*	*−43*
Expression of Affection	78	−08	*57*	*63*	*58*	*44*	*39*	29	18	16	10	06	*−51*	*−49*	*−56*	*−44*	*−62*	*−60*	*−56*
Positive Evaluation of Child	84	−07	*60*	*59*	*56*	*40*	*37*	29	12	13	06	−12	*−63*	*−51*	*−58*	*−44*	*−68*	*−62*	*−58*
Emotional Involvement	86	−25	35	*44*	*44*	*40*	*37*	28	*45*	*48*	20	11	−32	*−40*	−33	−29	*−46*	*−44*	*−51*
Achievement Demand	85	−14	22	20	17	14	10	08	31	*46*	12	−09	−29	−26	−20	−09	−30	−21	−28
Fostering Dependency	72	−11	06	16	21	13	22	01	24	24	18	13	−02	−15	−06	−05	−18	−24	−19
Excessive Contact	81	−30	−04	−04	06	01	18	24	10	18	30	06	−04	−05	05	06	−09	−09	−16
Intrusiveness	77	−24	−31	−32	−14	−11	07	04	21	20	34	23	23	08	26	27	18	10	12
Social Isolation of Child	80	−08	−19	*−36*	−22	−26	−11	−04	−02	−10	12	−15	07	17	25	26	07	14	15
Wish to Control Child	88	−18	*−50*	*−49*	*−36*	−24	−06	−06	09	32	33	16	*40*	32	*48*	*42*	*38*	*38*	27
Strictness	77	04	*−52*	*−44*	*−41*	−33	−24	−22	−01	16	12	00	*39*	*38*	*52*	*48*	*44*	*52*	*40*
Use of Fear to Control Child	85	−10	*−67*	*−61*	*−54*	*−36*	−18	−20	−02	07	15	21	*63*	*48*	*64*	*55*	*63*	*66*	*57*
Perceives Child as Burden	81	19	−34	*−37*	*−48*	*−42*	−14	−35	−06	*−46*	−06	−09	01	16	25	27	19	14	26
Irritability	89	−02	*−59*	*−57*	*−47*	−30	−12	−27	06	10	15	29	*64*	*40*	*54*	*38*	*63*	*53*	*51*
Punitiveness	86	−02	*−54*	*−55*	*−45*	−31	−12	−22	−04	01	15	21	*53*	28	*46*	28	*54*	*42*	*42*
Ignoring	89	15	*−63*	*−70*	*−60*	*−49*	*−38*	*−38*	−14	−22	−12	08	*63*	*47*	*60*	*43*	*68*	*64*	*62*

Note.—Decimal points omitted; $N = 31$.

* Significant correlations are italicized; correlations above .36 are significant at .05 level and above .47 are significant at .01 level.

TABLE 2

CORRELATIONS OF CLUSTER SCORES FOR ZERO- TO THREE-YEAR OBSERVATIONS AND NINE- TO FOURTEEN-YEAR INTERVIEWS

(Combined interrater reliabilities in diagonal cells)

Cluster	1	2	3	4
1. Love–Hostility (0–3 years)	(94)			
2. Love–Hostility (9–14 years)	68*	(91)		
3. Autonomy–Control (0–3 years)	−14	−12	(87)	
4. Autonomy–Control (9–14 years)	09	34	26	(84)

Note.—Decimal points omitted; $N = 31$.

* Significant beyond the .001 level of significance using a two-tailed test.

TABLE 3

CORRELATIONS BETWEEN ADDITIONAL ZERO- TO THREE-YEAR MATERNAL RATINGS AND NINE- TO FOURTEEN-YEAR RATINGS OF POSITIVENESS OF MOTHER-CHILD INTERACTION

0- to 3-Year Maternal Behavior Rating	Combined Interrater Reliability	Correlation with 9- to 14-Year Rating of Positiveness of Mother-Child Interaction (Reliability = 88)
Cooperativeness	85	58**
Rejection of the Homemaking Role	83	−45
Poor Physical Health	81	−40
Negative Emotional State	70	−40
Subjective Financial Stress	88	−36
Estimated Intelligence	85	36*
Dominance	84	−24
Mood Swings	79	−24
Self-Abasement	72	19
Overconscientiousness	80	13
Dependency	72	−10
Sociability	82	−04
Narcissism	73	01

Note.—Decimal points omitted; $N = 31$.

* Correlations above .36 significant at .05 level using two-tailed test of significance.

** Correlations above .47 significant at .01 level using two-tailed test of significance.

of cooperativeness of the mother and rejection of the homemaking role with the later rating of positiveness of mother-child interaction, reported in Table 3, can be attributed to the fact that these variables have high loadings on the dimension of Love versus Hostility that is also consistent across this period. Poor physical health of the mother, negative emotional states of the mother, and financial stress are significantly negatively correlated with later positive mother-child interaction. The finding of a positive correlation between estimated intelligence of the mother and later positiveness of mother-child interaction may reflect a halo effect in rating either maternal intelligence or behavior or may reflect a true relationship.

DISCUSSION

The correlation of .68 between the scores of these *S*s for the two sets of data for the dimension of Love versus Hostility, although significantly consistent, may not appear as high as one might wish in order to make inferences about maternal behavior from one age level to another. However, when we consider the many factors that might reduce the correlation between the two sets of data, the size of the correlation is relatively high. There are probably true changes in the behaviors of individual mothers during the 6- to 11-year interval between the observations and the interviews. The two different methods of collecting data, interview and observation, might lead to different impressions of characteristic behavior. The different situations in which the data were collected, the university research setting and the child's home, might influence the behavior of the mothers in different ways. Similar factors may also have contributed to the insignificant correlation between age levels for the Autonomy versus Control dimension. If we assume that the ratings of the Autonomy versus Control dimension are as valid as those for the Love versus Hostility dimension, it would appear that there are greater changes in maternal behavior on this dimension through time. One interpretation of this difference might be that a child's need for love and a positive relationship with the mother is consistent through time, while the child's need for autonomy changes from a need for total dependence and a close relationship at birth to a need for relatively complete autonomy as an adult. Some mothers who have a close relationship with the child described by the ratings of excessive contact, fostering dependency, low autonomy, and intrusiveness during infancy may shift to give an age-appropriate autonomy as the child matures. Other mothers who have a more distant relationship with the infant may become more possessive, intrusive, and controlling later as the child becomes more independent and more active. Other hypotheses might be that there are individual variations in the degree to which children seek autonomy or there may be changes in the mother's capacity to satisfy the child's needs, and that these changes are reflected in changing scores on this dimension.

The finding that early ratings of poor physical health of the mother, subjective financial stress, and negative emotional states (gloominess, dullness, unresponsiveness) have significant negative correlations with later positiveness of the mother-child-relationship suggests that environmental stresses felt by the mother, reduced physical vitality, or anything which contributes to a negative emotional adjustment of the mother may impair the development of a good relationship with the child. Preoccupation with other problems may reduce the capacity of the mother to provide warmth and security for the child, or these results may be interpreted by the frustration-aggression hypothesis (Dollard, Doob, Miller, Mowrer, & Sears, 1939). Of course, poor health, subjective financial stress, and negative emotional states may all reflect emotional maladjustment of the mother rather than objective stresses.

The findings of this study suggest the need for caution in the interpretation of retrospective studies of the relationship of maternal behavior to the development of children. The relatively low consistency for the dimension of Autonomy versus Control, the control end of which is similar to the concepts of overpossessiveness and overprotectiveness, suggests that it may be difficult to make inferences about early maternal behavior relevant to this dimension from interview data collected at a later date. Of course, the interview data of this study were concerned with current family behavior rather than with attempts to reconstruct what had happened earlier. Although the correlation between the Love versus Hostility scores for the two age levels provided support for making inferences about maternal behavior from one age level to another, the size of the correlation does not preclude, in individual cases, relatively large changes through time.

SUMMARY

An interest in the influence of maternal behavior upon development of the child suggested two hypotheses:

1. A mother's behavior with her infant is significantly related to her behavior with the same child at preadolescence.
2. Data on a mother's behavior with her child collected from an interview are significantly related to data collected from direct observation of mother-child interaction.

Observational data collected during the first three years of the child's life in testing sessions at the University of California Institute of Child Welfare and interview data collected in the home between the child's age of 9 to 14 years permitted a test that confounded the two hypotheses.

The finding of a significant correlation between scores on the two sets of data for a dimension of Love versus Hostility led to acceptance of both hypotheses for that dimension. The finding of an insignificant, although positive, correlation between scores for the dimension of Autonomy versus Control could not be definitely interpreted since this result might be due either to low validity for either the observation or interview data or both, or to true inconsistency in maternal behavior for this dimension. It was found that early ratings of poor physical health of the mother, subjective financial stress, and negative emotional states had significant negative correlations with later ratings of positiveness of mother-child interaction.

REFERENCES

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

GUTTMAN, L. A new approach to factor analysis: The Radex. In P. F. Lazarsfeld (Ed.), *Mathematical* *thinking in the social sciences*. Glencoe, Ill.: Free Press, 1954.

JONES, H. E., & BAYLEY, NANCY. The Berkeley growth study. *Child Develpm.*, 1941, **12**, 167–173.

SCHAEFER, E. S., BELL, R. Q., & BAYLEY, NANCY. Development of a maternal behavior research instrument. *J. genet. Psychol.*, 1959, **95**, 83–104.

SCHAEFER, E. S. A circumplex model for maternal behavior. *J. abnorm. soc. Psychol.*, 1959, **59**, 226–235.

THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. Chicago Press, 1947.

(Received July 14, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 7-14

INFANTILE TRAUMA, GENETIC FACTORS, AND ADULT TEMPERAMENT

GARDNER LINDZEY, DAVID T. LYKKEN, AND HARVEY D. WINSTON
University of Minnesota[1]

IN SPITE of a welter of empirical and theoretical activity centering upon the role of infantile experience as a determinant of adult personality it is evident that there are few important issues in this area that have been satisfactorily resolved. Prominent among these open issues is the potential contribution of gene structure to the relation between infantile experience and adult temperament. Although King (1958) has identified gene factors as one of seven major parameters to be considered in studying the relation between infantile experience and adult effects, there is little in the way of compelling empirical findings concerning the role of genetic factors in this setting.

This sparse activity in the world of observation exists in spite of a history of theoretical interest in this issue. There are few general theories of behavior that fail to take into consideration, at least implicitly, the potential impact of constitutional or genetic factors upon early experience and its consequences. Perhaps the most influential and persistent emphasis upon the role of heredity in this context has been provided by Freud (1905) with his discussions of the determinants of fixation, but his position differs from other personality theorists on this issue primarily in regard to explicitness.

The present study was designed to provide further evidence relevant to this general issue. Because of traditional difficulties involved in using human subjects for studies spanning the life history of the organism and the additional requirement that genetic variation be controlled, it was necessary to use subjects drawn from inbred strains of mice.

While the most unusual aspect of this study inquired into the interaction between gene factors and the effects of infantile trauma, it would be a rash investigator indeed who risked his all on an interaction term. Consistent with this premise is the fact that our study also provides information concerning the general effects of infantile trauma, their situational generality and duration, as well as further evidence regarding the influence of genetic factors upon various measures of temperament.

A word should be said concerning previous results that are relevant to the relation between infantile trauma and emotionality, for here, as well as in connection with the interaction effect we have just discussed, we have made specific predictions prior to the study. Clear evidence of an increase in emotionality in mice as a result of infantile trauma has been demonstrated by Hall and Whiteman (1951), while other investigators working with mice (Stanley & Monkman, 1956) and with rats (Griffiths & Stringer, 1952) have failed to find any significant differences in emotionality between control and traumatized animals. One investigator (Ader, 1957) actually has reported results indicating that treated rats were *less* emotional than control animals. Although these studies have led to conflicting results, the firmness of the findings of Hall and Whiteman (1951), coupled with consistent evidence from human investigations (Bowlby, 1951) and a boost from psychoanalytic theory, led us to predict that early traumatic experience would produce an increase in emotionality (timidity).

In summary, this study was designed to provide information concerning:

1. The importance of genetic factors as determinants of emotionality, timidity, and activity.

2. The general effects of infantile trauma, their situational generality, and the duration of these effects through time.

3. The influence of genetic factors upon the relation between infantile trauma and adult temperament.

---

[1] From the Center for Personality Research. This investigation was supported by a grant from the Ford Foundation. We are deeply grateful to John J. Bittner for his generous assistance and detailed advice concerning the care and breeding of mice. Mary Ahlquist, Samuel Berman, and Arthur Hill all rendered valuable aid at various stages in this study.

METHOD

Subjects

The (*Ss*) of this investigation were offspring of breeding animals belonging to four different inbred strains of mice. The parent animals were obtained from a colony maintained by John J. Bittner of the Cancer Biology Laboratories of the University of Minnesota and represent the outcome of many generations of brother-sister matings. The conventional labels for these strains and the approximate number of generations of controlled brother-sister matings are as follows: C57BL/1 or B/1 (51 generations); C3H (93 generations); DBA/8 or D/8 (27 generations); JK (71 generations). The first three strains were selected on the basis of prior information (Lindzey, 1951) indicating diversity in emotionality, while the JK strain was selected because observation of them in the colony suggested that their behavior was quite different from the other strains. They were also selected so that all four strains could be identified easily by their distinctive coat color.

Within each strain the successive litters at birth were assigned alternately to either the experimental or the control group with the restriction that when the second litter by a particular mother was born that it would be given a different assignment than the first litter. This shifting procedure was continued through all litters until the desired number of animals in either the control or experimental group was achieved, and all subsequent litters were then assigned to the other group until the necessary number was reached. Because of the impossibility of predicting survival rates precisely for a period of almost four months, a variable number of animals resulted in the different experimental groups. The smallest number in any single group was 23 while the largest number was 36. The total number of animals included in the study was 259.

Each litter was raised in a separate cage with the mother until 24 days of age, at which time the young were weaned and segregated by sex. Animals were individually identified by an ear punch code, but this operation was not carried out until completion of the first consequent measure in order to avoid the possibility that this relatively traumatic procedure would wipe out the effects of our experimental treatment.

Infantile Trauma

Control and experimental animals were raised under the same conditions except that at four days ot age those animals assigned to the experimental group were on four successive days exposed to an extremely loud, high frequency, auditory stimulus. The procedure followed was essentially the same as that described by Hall and Whiteman (1953) and involved placing the mice on a thin cardboard container in a #1 12-gal. wash tub with a doorbell (#504 Eclipse) fixed to the side. After a 2-min. interval the doorbell was rung for 2 min., following which the mice were left in the tub for another two minutes before being returned to their cages. On the first three trials all animals were placed together in the tub, but on the fourth trial they were separated. The procedure for the control animals was precisely the same except that the bell was not rung while they were in the tub.

Consequent Measures

30-Day Open Field Test. Beginning at 30 days of age each mouse was placed for 2 min. in the same type of wash tub in which the auditory trauma had been administered. The tub was brightly lighted by a shielded 150-w. glazed bulb that was placed directly over the tub at a height of approximately $1\frac{1}{2}$ ft. The mouse was left in the tub for 2 min. at the end of which time the incidence of defecation and urination was recorded and the mouse returned to his living cage. This measure was repeated on 10 successive days. The measure of emotionality was simply presence or absence of defecation and/or urination during each trial.

Stovepipe Test. This instrument was devised by Stone (1929) as a measure of timidity or of the "dominance of hiding tendency over hunger," (p. 36) and was used by Hall and Whiteman (1951) in the study referred to earlier. In this measure the mouse was placed in a starting box connected with a U-shaped stovepipe (each unit approximately 2 ft. in length) that led to a goal box containing a food receptacle holding wet mash. Thus, all that the mouse had to do to reach the food was to enter the stovepipe, follow the passage, and leave the dark pipe for the more brightly illuminated goal box. The mice were again given 10 trials on successive days, commencing at 70 days of age, and they were run under $22\frac{1}{2}$ hrs. of food deprivation. The measure employed was the total time that it took on the 10 trials for the mouse to enter the goal box. The few mice that on particular trials failed to leave the starting box within 5 min. were on these trials arbitrarily assigned the same time as the mouse that took the longest time in reaching the goal box.

100-Day Open Field Test. Commencing at 100 days of age the open field test was repeated for 10 trials with the measure of emotionality consisting of presence or absence of defecation and/or urination summed across the 10 trials for each *S*. In addition a measure of motility was secured by dividing the wash tub into 12 spaces of equal area and counting the number of spaces that each mouse moved through during the two minute trial and again summing the results of the individual trials over the 10 days.

Analysis of Data

The obvious method of statistical analysis for our data was a 2 × 4 analysis of variance and this technique was used for all data except those derived from the first open field test. In the latter case the absence of individual identification made it impossible to summate scores over the 10 trials and, in general, made it virtually impossible to devise a sensitive measure of the interaction hypothesis. Demonstration of strain differences and treatment effects was not particularly difficult, in part because they were relatively evident, and here we have relied upon descriptive presentation and the use of such simple devices as the sign test. The best estimate of interaction that could be devised, however, was to adapt the chi square test and then to apply this test separately to the data for each of the 10 days. This is obviously a somewhat unsatisfactory test of the interaction. We have not hesitated to apply one-tailed tests of significance to data related to directional hypotheses stated in advance of the study.

For purposes of analysis of variance we reduced all groups to the size of the smallest group in order to avoid the complexities of disproportionality. *Ss* were eliminated on a chronological basis beginning with the animals that were last to complete the procedures. In the case of descriptive presentation of results we have included all animals studied with an appropriate indication of the number in each group.

RESULTS AND DISCUSSION

Gene Determinants of Temperament

Here we are concerned with strain differences in two open field measures of emotionality, the stovepipe measure of timidity, and an open field measure of motility. The findings are clear and dramatic. No sentient person could observe the data reported in Tables 1–6 and conclude for any one of these attributes that the animals in all the strains could have been drawn from a single common population. In some instances the strains are so different as to show virtually no overlap. In view of the fact that the only known difference between these strains lies in their genetic constitution, it seems a prudent conclusion that genetic factors make a considerable contribution to each of our three measures of temperament.

Our findings are given added weight when coupled with the results of a number of earlier investigations using both rats (Hall, 1938; Broadhurst, 1958) and mice (Lindzey, 1951; Thompson, 1953, 1956; McClearn, 1959), all of which demonstrate wide strain differences on comparable measures. More interesting than this general congruence is a specific examination of the incidence of defecation on the part of the three strains included in both the present study and the previous investigation by Lindzey (1951). In each study, C57, DBA, and C3H strains were employed and in each case an open field test of emotionality was utilized. In Table 2 the percentage of animals defecating in each strain in the earlier study, and the percentage of defecation over 10 trials in each strain in the present study, are compared. The figures are highly constant in spite of the fact that the data were collected with more than a decade of time intervening, with considerable variation in the details of the open field test, with different persons collecting the data, and with animals that are not even known to belong to the same subline within the various strains. Even worse, we *do* know that the two samples were drawn from colonies that have been separated long enough to provide ample opportunity for "genetic drift" as well as spontaneous mutation. Given all of these sources of potential variation it is quite surprising to find the consistency of the strains and the measure so great. Thus, we are afforded the luxury of simultaneously gaining confidence in the stability of the organism we are studying and the reliability of our measure.

In general, these findings serve to emphasize again the pervasive impact upon behavior of genetic factors and thus strongly suggest the inadequacy of attempts to formulate or ac-

TABLE 1

PERCENTAGE OF ANIMALS DEFECATING DAILY ON 30-DAY OPEN FIELD TEST

Strain	Treatment	*N*	Day										Average
			1	2	3	4	5	6	7	8	9	10	
C57BL/1	Control	28	28.6	28.6	32.1	53.6	35.7	39.3	46.4	32.1	28.6	17.9	34.3
C57BL/1	Experimental	25	32.0	36.0	40.0	40.0	60.0	52.0	36.0	40.0	36.0	32.0	40.4
C3H	Control	34	85.3	88.2	82.3	64.7	73.5	70.6	85.3	73.5	73.5	73.5	77.1
C3H	Experimental	39	92.3	92.3	92.3	79.5	82.0	84.6	92.3	94.9	94.9	100.0	90.5
DBA/8	Control	37	64.9	81.1	73.0	70.3	62.2	64.9	62.2	64.9	59.5	67.6	67.0
DBA/8	Experimental	35	54.3	80.0	82.9	80.0	82.9	82.9	74.3	57.1	77.1	71.4	74.3
JK	Control	29	58.6	79.3	82.8	86.2	100.0	89.6	96.5	72.4	93.1	86.2	84.5
JK	Experimental	32	90.6	87.5	84.4	90.6	93.8	96.9	90.6	87.5	87.5	93.8	90.3

TABLE 2

CONSISTENCY OF EMOTIONALITY MEASURE IN 1951 STUDY AND PRESENT STUDY

	C57BL	C3H	DBA
Percent defecation (1951)	37 (N = 100)	72 (N = 100)	68 (N = 100)
Percent defecation (present study)	34 (N = 28)	77 (N = 34)	67 (N = 37)

TABLE 3

ANALYSIS OF VARIANCE OF STOVEPIPE TEST

Source	*df*	Sum of Squares	Mean Square	*F*	*P*
Treatment	1	2,350,782	2,350,782	1.93	<.10*
Strain	3	241,013,412	80,337,804	65.96	<.0001
Interaction	3	10,241,048	3,413,682	2.80	<.05
Within (error)	176	214,356,579	1,217,935		
Total	183	467,961,821			

* One-tailed test.

count for behavior that do not direct careful attention to the role of hereditary factors. It is interesting to note that it is not necessary to inbreed selectively for particular characteristics in order to produce clear evidence for the influence of genic factors upon behavior. Inbreeding of the present animals was uninfluenced by the temperament characteristics of interest in the present study, and, in spite of this, we find the particular genetic patterns that have been isolated are dramatically different in their behavioral consequences.

Influence of Infantile Trauma upon Temperament

What exactly were the effects of the auditory, infantile trauma? The reader will recall our prediction that this experience would lead to an increase of emotionality and timidity. Examination of the data summarized in Table 1 reveals that in spite of the awkwardness resulting from our inability to identify individual animals, there is little doubt that the experimental treatment led to an increase in emotionality as measured by the 30-day test. Inspection of the proportion of experimental and control animals defecating on each of the 10 trials within the four strains reveals that 32 of the 40 comparisons point to the greater emotionality of the traumatized animals. This incidence of "hits" is well beyond the 1% level of significance, even utilizing the relatively insensitive sign test. Actually, the within-strain results for the C3H animals (on all 10 trials the experimental animals showed a greater proportional defecation) are by themselves significant at below the 5% level using the sign test, while results for the other strains (7 or 8 trials out of 10 in which experimentals defecate more) approach significance.

The results of the stovepipe test, summarized in Table 3, provide suggestive evidence ($p < .10$) for a treatment effect indicating greater timidity (slower in getting to reward) on the part of the traumatized mice. Consistently, we find in Table 4 that at the 100-day open field test there is also a difference in emotionality indicating that the treated animales were more emotional. Only in the case of motility, as revealed in Table 5, do we find little indication of treatment effects upon behavior.

To summarize, there seems little doubt that subjecting infant mice to a noxious, auditory stimulus of high intensity leads to an increase in emotionality in adulthood and there is some basis for believing that there is also an increase in timidity. There is no evidence, however,

TABLE 4

ANALYSIS OF VARIANCE OF 100-DAY OPEN FIELD TEST: EMOTIONALITY

Source	*df*	Sum of Squares	Mean Square	*F*	*P*
Treatment	1	16.5093	16.5093	2.86629	<.05*
Strain	3	679.4458	226.4819	39.3211	<.0001
Interaction	3	2.1082	.7027	—	—
Within (error)	176	1013.7287	5.7598		
Total	183	1711.7920			

* One-tailed test.

TABLE 5

ANALYSIS OF VARIANCE OF 100-DAY OPEN FIELD TEST: MOTILITY

Source	*df*	Sum of Squares	Mean Square	*F*	*P*
Treatment	1	14,635.2	14,635.2	1.29	<.50
Strain	3	1,421,619.4	473,873.1	41.8	<.0001
Interaction	3	24,268.4	8,089.5	.071	—
Within (error)	176	1,994,390.6	11,331.8		
Total	183	3,454,213.6			

TABLE 6
MEANS AND *SD*s FOR STOVEPIPE TEST AND 100-DAY OPEN FIELD TEST

Strain	Treatment	Stovepipe		100-Day Emotionality		100-Day Motility	
		Mean ($N = 24$)	*SD*	Mean ($N = 23$)	*SD*	Mean ($N = 23$)	*SD*
C57BL/1	Control	2301	1339	4.26	3.40	401.2	135.3
C57BL/1	Experimental	2069	882	4.57	3.19	355.0	165.1
C3H	Control	3322	1669	8.43	1.58	198.2	110.7
C3H	Experimental	4312	1039	9.35	1.05	188.1	85.9
DBA/8	Control	773	436	6.39	2.84	272.5	63.3
DBA/8	Experimental	778	407	6.96	2.71	287.7	99.7
JK	Control	1480	1061	8.87	1.51	163.3	67.1
JK	Experimental	1601	1007	9.48	1.02	133.0	53.7

that this experience influences motility in an open field setting.

These findings confirm our prior expectations and fit very neatly with the results reported by Hall and Whiteman (1951) on the basis of similar experimental operations. However, our results directly contradict the conclusion of Ader (1957), based upon an investigation dealing with rats, and are inconsistent with the findings of Griffiths and Stringer (1952), and Hunt and Otis (1955) based on observation of rats, as well as the findings of Stanley and Monkman (1956) derived from the study of mice. There are so many parametric differences between all of these studies (species of animal, nature of trauma, age at trauma, age at consequent measure) that it is probably useless to attempt a rational analysis of the basis for these experimental differences. It is worth note, however, that there is actually no evidence in any study for infantile trauma *decreasing* emotionality. Although Ader has suggested this possibility, his own data, in the only study in which he employed stimuli that could reasonably be considered noxious or traumatic (Ader, 1959), suggest a heightening of emotionality as a result of infantile trauma. Moreover, Ader's attempts to account for the findings of Hall and Whiteman on the basis of the fact that they used seizure-susceptible mice and an infantile trauma identical to that used in producing audiogenic seizures, is not supportable in view of the fact that the present study led to comparable results with strains such as the C57BL/1 and C3H, which are very low in seizure incidence (Lindzey, 1951).

Duration of traumatic effects. Because of the imprecision of our measures and the differences in the form of our data at age 30 days and age 100 days, it is impossible to make any exact estimate of the extent to which the experimental effects may have diminished with the passage of time. However, it is evident that at 30 days, at 70 days, and at 100 days, the influence of the infantile trauma is clearly observable. Whatever may have been lost in intensity, the main effect is still manifest at the time of the last measure. Actually, a close descriptive inspection of the results of the present study, and the results contained in the Hall and Whiteman paper, suggest that there may have been some diminishing of treatment effects by the time of the 100-day test, but there is no firm evidence to support this contention.

Generality of traumatic effects. One obvious and important question has to do with the narrowness with which the effects of the infantile trauma are linked to situations resembling the original traumatic setting. Obviously, the implications of the trauma for the organism are very different if it leads to a generalized response disposition that is elicited in many settings, rather than leading to effects that can be observed only in situations very similar to the original one. Both Beach and Jaynes (1954) and Ader (1959) attempt to account for the increased emotionality following infantile trauma reported by Hall and Whiteman on the basis of the similarity between the traumatic setting and the setting in which the emotionality was measured. The results of the present study are quite specific in refuting this proposal, as we secured consistent empirical findings in connection with both Stone's measure of timidity and Hall's open field measure. The stimulus differences between the stovepipe test and the open field test are so profound that if we are willing to concede the possibility of stimulus generalization from the original setting to the stovepipe test, we have already granted tremendous generality to the effects of the trauma. In brief, then, our findings suggest that the results of the infantile trauma are of a relatively general nature and should be discernible in a wide variety of different settings.

Impact of Genetic Variation upon Effects of Infantile Trauma

We have already seen that our behavioral measures of temperament are heavily influenced by genetic variation and that these same measures covary with traumatic, infantile experience. The essential question remaining is whether these two sets of determinants show any interaction. Are the changes in temperament that can be attributed to early infantile experience in part dependent upon the gene structure of the organism undergoing the experience?

It was impossible to devise a completely satisfactory test of the interaction effect for the 30-day open field test of emotionality because the individual animals were not identified. Consequently, the best we could do was to measure the interaction between strain and treatment effects for each of the 10 trials individually. The resulting χ^2 coefficients provide no evidence to support the hypothesis of interaction. It is true that we would have preferred to use a single more sensitive measure of emotionality (incidence of defecation summed over 10 days), and it is possible to speculate that with this increased sensitivity we would have been able to detect an interaction. There is, however, no direct evidence to support this contention.

Our second measure of interaction between strain and treatment effects deals with the stovepipe measure of timidity and, as the results summarized in Table 3 indicate, we found clear evidence for the existence of an interaction. These results suggest that changes in timidity resulting from the infantile trauma are in part dependent upon the strain (genetic makeup) of the mouse. In particular, the C57BL/1 mice seemed much less influenced by the experimental treatment than the other strains. The third test of the interaction hypothesis is presented in Tables 4 and 5, where we find no evidence for an interaction between strain and treatment in effects upon emotionality or motility at the 100-day test.

The simplest conclusion that can be derived from these findings is that there is an interaction between gene factors and infantile trauma for timidity, as measured by the stove pipe test, but not for emotionality or motility. While this finding may ultimately be substantiated, there are certain unsatisfactory aspects to our test of the interaction of strain and infantile trauma in the case of the two emotionality measures and the measure of motility. The test for the 30-day data had to be carried out upon single observations rather than observations cumulated over 10 days and the test of interaction for the 100-day data may have suffered from a diminishing of treatment effects at this stage. It is known that the analysis of variance measure of interaction is not a powerful test and it is possible that interaction could be demonstrated only with more powerful treatment effects or a larger number of *S*s. Our failure to find treatment effects upon motility made our interaction measure in this area relatively meaningless.

So far as our hypothesis is concerned, we have obtained evidence for the existence of interaction between strain and infantile treatment effects as predicted, but this relationship was not so general as we had expected. Our conviction concerning the existence of some degree of interaction is strengthened by the results obtained by King (1959), demonstrating an interaction between the effects of infantile handling and membership in different subspecies, and Valenstein, Riss, and Young (1955), suggesting a relationship between the effects of isolation upon sexual behavior and membership in a particular inbred strain. An additional study by King (1957) also provides evidence suggestive of such an interaction, although the experimental treatment (isolation) did not take place in this case until after weaning. The findings indicate that changes in aggressive behavior that were produced by isolation in one mouse strain were not duplicated in a second strain. There is unfortunately, no direct test of the difference between the two strains in treatment effects. Joint consideration of all of these findings provides a relatively firm basis for concluding that the consequences of infantile trauma are in part dependent upon the genetic structure of the organism experiencing the trauma.

What are the implications of such a finding for psychological theory? One may contend that these results, even secured in connection with mere mice or guinea pigs, provide a type of confirmation of Freud's assertion that any attempt to map early experience into adult behavior must allow for the contribution of genetic factors. While such a statement has a rather hollow sound at the level of human behavior because of the little that is known concerning the gene structure of man, it does have some specific empirical implications. For example, our results would suggest that when an investigator works with *S*s of unknown or uncontrolled heredity, it is altogether possible to conduct an otherwise exemplary study of the effects of infantile experience and fail to find evidence for such effects. Or, more generally, we may expect that a variety of different empirical findings might be observed in similar studies as a consequence of investigators dealing with *S*s of various genetic backgrounds, rather than as a result of faulty experimental technique. In general, this is a finding that makes life more complex for both investigator and theorist, and in an area where there has never been any shortage of complexity.

SUMMARY

This investigation was concerned with the effects of infantile trauma upon adult temperament, the influence of genetic factors upon temperament, and the possibility of an interaction between early experience and gene structure in their influence upon adult behavior.

Four strains of homozygous mice were used in the study, with the infant offspring assigned by litter to either an experimental or control group. In all, 259 mice were studied. The experimental mice, beginning at four days of age, were exposed on four successive days to a noxious, auditory stimulus while the control animals were treated in an identical manner except that they were not exposed to the traumatic stimulus. At 30 days of age all mice were examined for 10 successive days in an open field test of emotionality and beginning at 70 days of age the mice were again observed for 10 successive days in a stovepipe test of timidity. A final measure of emotionality, and a measure of motility, were secured from 10 days of observation in an open field test beginning at 100 days of age.

The data obtained provide clear and compelling evidence for the importance of genetic factors as determinants of emotionality, timidity, and motility. The four strains displayed marked differences in all three attributes. There was also direct evidence for the influence of infantile trauma upon emotionality and suggestive evidence in regard to timidity, but no evidence for such an influence upon motility. The effects of the infantile trauma were enduring, extending at least to an age of 100 days, and were not limited to stimulus situations closely similar to the orginal traumatic situation. We found evidence of an interaction between the effects of infantile trauma upon stovepipe timidity and genetic factors, but there was no evidence for such an interaction in the case of emotionality. All of our positive findings are supported by evidence supplied by other investigators.

These findings not only demonstrate the central developmental importance of genetic factors and infantile trauma, they also underline the relative complexity of the relationship between infantile experience and adult behavior. Given an interaction between gene factors and infantile trauma it is readily understandable that inconsistent results might be observed by investigators working with heterozygous *S*s of unknown gene structure. Finally, we have pointed to the consistency between the present findings and Freud's formulations concerning the role of constitutional factors in the developmental process.

REFERENCES

ADER, R. Effects of early experience on emotionality. *Amer. Psychologist*, 1957, **12**, 410.

ADER, R. The effects of early experience on subsequent emotionality and resistance to stress. *Psychol. Monogr.*, 1959, **73**(2, Whole No. 472).

BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, **51**, 239-264.

BOWLBY, J. *Maternal care and mental health.* Geneva: World Health Organization, 1951.

BROADHURST, P. L. Determinants of emotionality in the rat: III. Strain differences. *J. comp. physiol. Psychol.*, 1958, **51**, 55–59.

FREUD, S. Three essays on the theory of sexuality. (Originally published 1905.) In *The standard edition of the complete psychological works of Sigmund Freud.* Vol. 7. London: Hogarth, 1953, Pp. 125–248.

GRIFFITHS, W. J. JR., & STRINGER, W. F. The effects of intense stimulation experienced during infancy on adult behavior in the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 301–306.

HALL, C. S. The inheritance of emotionality. *Sigma Xi Quart.*, 1938, **26**, 17–27.

HALL, C. S., & WHITEMAN, P. H. The effects of infantile stimulation upon later emotional stability in the mouse. *J. comp. physiol. Psychol.*, 1951, **44**, 61–66.

HUNT, H. F., & OTIS, L. S. Restricted experience and "timidity" in the rat. *Amer. Psychologist*, 1955, **19**, 432.

KING, J. A. Relationships between early social experience and adult aggressive behavior in inbred mice. *J. genet. Psychol.*, 1957, **90**, 151–166.

KING, J. A. Parameters relevant to determining the effect of early experience upon the adult behavior of animals. *Psychol. Bull.*, 1958, **55**, 46–58.

KING, J. A., & ELEFTHERIOU, B. E. Effects of early handling upon adult behavior in two subspecies of deermice, *Peromyscus maniculates. J. comp. physiol. Psychol.*, 1959, **52**, 82–88.

LINDZEY, G. Emotionality and audiogenic seizure susceptibility in five inbred strains of mice. *J. comp. physiol. Psychol.*, 1951, **44**, 389–393.

MCCLEARN, G. E. The genetics of mouse behavior in novel situations. *J. comp. physiol. Psychol.*, 1959, **52**, 62–67.

STANLEY, W. C., & MONKMAN, J. A. A test for specific and general behavioral effects of infantile stimulation with shock in the mouse. *J. abnorm. soc. Psychol.*, 1956, **53**, 19–22.

STONE, C. P. Wildness and savageness in rats of different strains. In K. S. Lashley (Ed.), *Studies in the dynamics of behavior.* Chicago: Univer. Chicago Press, 1932. Pp. 3–55.

THOMPSON, W. R. The inheritance of behavior: Behavior differences in fifteen mouse strains. *Canad. J. Psychol.*, 1953, **7**, 145–155.

THOMPSON, W. R. The inheritance of behavior: Activity differences in five inbred mouse strains. *J. Hered.*, 1956, **47**, 147–148.

VALENSTEIN, E. S., RISS, W., & YOUNG, W. C. Experiential and genetic factors in the organization of sexual behavior in male guinea pigs. *J. comp. physiol. Psychol.*, 1955, **48**, 397–403.

(Received May 25, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 15-20

OPINION CHANGE AS A FUNCTION OF EXTERNAL COMMITMENT AND AMOUNT OF DISCREPANCY FROM THE OPINION OF ANOTHER[1]

MILTON E. ROSENBAUM AND DOUGLAS E. FRANC
State University of Iowa

A CONSIDERABLE amount of research evidence (Hovland, Janis, & Kelley, 1953; Osgood & Tannenbaum, 1955) has indicated that an attempt to influence change of the opinions of an audience is facilitated by a positive evaluation of the communicator. One strategy that a communicator may employ to gain a favorable appraisal is to present his conception of the audience's current position on a relevant dimension of opinion. This technique is frequently seen as setting the stage for acceptance of the communicator as a discerning and straightforward person (Yeager, 1951) or as a means of establishing a tie to the audience (Powell, 1951). On some occasions, such attribution of position may interfere with the communicator's intention rather than furthering it. The purpose of the present study is to investigate some of the conditions under which attribution of position affects the response to an attempt to change opinions.

In an earlier study (Rosenbaum & Zimmerman, 1959) dealing with an attempt to induce change in the opinions of a sample of students in a Southern university concerning segregation, it was found that attribution of a position that was congruent with the previously recorded opinions of the *S*s resulted in significantly greater resistance to a persuasive communication in comparison with the change of opinion found among control *S*s to whom no attribution was made. No significant effect was found for an incongruent attribution. The postitive finding was interpreted as involving a process tentatively labeled *external commitment*. The term "external commitment" is chosen on analogy to the conventional variable of *self-commitment*. Self-commitment is typically evoked by having the person indicate publicly an opinion or judgment prior to exposure to some procedure intended to influence his views. Results of several studies (Deutsch & Gerard, 1955; Fisher, Rubinstein, & Freeman, 1956; Hovland, Campbell, & Brock, 1957) indicate that resistance to the succeeding influence attempt accompanies this initial verbalization of position.

External commitment can be hypothesized to function similarly to self-commitment. It too should produce resistance to influence attempts when the attribution is *congruent* with the person's current response tendency and opposite to the direction of the influence attempt. External commitment procedures also allow attribution of a position that is *incongruent* with an individual's current response and in this case, susceptibility to an attempt to change opinion should be facilitated. In the present study, conditions are arranged by means of which congruent and incongruent external commitment procedures can be compared with a control or no commitment condition.

Evaluation of the earlier study (Rosenbaum & Zimmerman, 1959) led to the hypothesis that the failure to demonstrate the facilitative effect of incongruent commitment may have been due to the deep-seated and emotional character of the opinion area studied or to the extreme amount of change advocated or both. In the present study, varying amounts of discrepancy between *S*'s opinion and the opinion of the influencing agent are employed in order to deal directly with the second of the suggested factors. Apparently less deep-seated opinions are studied but no attempt is made to manipulate this variable directly.

The relationship between amount of opinion discrepancy and amount of change obtained has been studied by Goldberg (1954) and Hovland and Pritzker (1957) with the results of their research indicating a positive relationship; the greater the opinion discrepancy, the

---

[1] The authors wish to express their gratitude to Harold P. Bechtoldt and Dee W. Norton for statistical and editorial assistance.

greater the amount of change obtained.[2] A portion of the present study can be considered a replication of Hovland and Pritzker's since similar items and influence procedures are employed. Hovland and Pritzker recognize with appropriate caution[3] that the positive relationship obtained may apply only to certain types of opinion items and up to certain degrees of opinion discrepancy. Because of the possibility that the inhibiting or facilitative effect of external commitment is contingent on the amount of opinion discrepancy in the persuasive attempt, a factorial design is used by means of which potential interactive effects can be evaluated.

METHOD

Approximately 400 students in an introductory psychology class responded to a questionnaire that contained, among other items, adaptations of the 12 opinion items employed by Hovland and Pritzker (1957). A sample item reads "All things considered, Washington was a greater president than Lincoln." The questionnaire required response on a six-point scale ranging from "strongly agree," indicated by a +3, to "strongly disagree" which was represented by a −3. From this larger population, 72 *S*s "volunteered" to participate in the experimental session as part of their required hours of extra-class experimental participation.

The experiment was presented to the *S*s in individual sessions as being concerned with social perception and they were instructed to write a one-page essay on the topic "The Value of a College Education." Upon completion of the essay, *E* took it from the room informing *S* that the essay was to be read by a professional person, who would, on the basis of his reading, make judgments as to the *S*'s opinion on other topics. *S* was told that the professional people who were participating in the experiment "are highly competent in their own fields and have wide knowledge in other fields" and that the major focus of the study was to evaluate the social perceptiveness of people who have to deal with others.

A brief delay ensued during which *E* was ostensibly dealing with the professional person. *E* then returned with a booklet containing 24 pages. Each page had at the top the statement of an issue and below that three rows of six boxes. Each box represented a point on the agree-disagree continuum. The *S*s were specifically informed on a face sheet and via oral instruction that the first line of boxes on each sheet represented the professional person's judgment of *S*'s opinion on the particular issue at the top of the sheet. The second line of boxes on each sheet represented the professional person's own position and the third line was to be used by the *S* to indicate his own current opinion. Twelve pages of the booklet contained the critical items that had appeared in the earlier questionnaire and 12 pages presented filler items. A standard order of presentation was employed that permitted no more than two critical items to appear together.

Employing this form, the top line was used in order to administer the external commitment variable. The second line served for the opinion discrepancy variable and the third line represented the dependent responses to be evaluated along with the prequestionnaire responses.

The commitment variable was carried out in the following manner. One-third of the *S*s were placed in a condition of congruent commitment in which all 12 critical items were checked on the first line at the position these *S*s had indicated in their prequestionnaires. One-third of the *S*s were assigned to a condition of incongruent commitment in which check marks were placed three boxes away from the *S*'s prequestionnaire position thereby clearly attributing to him a position opposed to his original position. For example, if *S* had intially strongly disagreed with an item, he was attributed to a "slightly agree" position. The remaining one-third of the *S*s were exposed to a control condition in which no check marks appeared on the first line. These *S*s were told that the professional person's predictions were made on separate sheets and the presence of the top line on their sheets was to facilitate later scoring procedures.

The procedures to vary opinion discrepancy were carried out on the second line. One-third of the *S*s in each of the commitment conditions were exposed to a check mark one box away from *S*'s original position, one-third to a two-unit difference and the remaining third to three units. In all cases, the check mark on the second line was placed in a direction toward the opposite side of the agree-disagree dichotomy from that indicated by *S*'s original position. The 12 filler items were checked in a standard manner that precluded the repetition of similar check patterns throughout the booklet.

The above procedures result in a 3 × 3 factorial design with eight *S*s per cell. *S*s were assigned randomly to the resulting nine combinations of procedures manipulating commitment and opinion discrepancy.

The dependent variable, amount of change, was derived by relating the *S*'s response on the third line of the critical booklet items to his prequestionnaire responses to the same item. The six-point Likert scale was given unit values ranging from a score of one for "strongly agree" to a score of six for "strongly disagree". A difference score for an item was obtained by subtraction of the final score from the prequestionnaire score. The sign of the change was computed by assigning a plus to changes in the direction of the professional person's own opinion and a minus for changes away from the professional person's position.

---

[2] The use of the term "advocated change" in Hovland and Pritzker's (1957) study and others using similar operations is not completely consistent with the operations employed. In these studies, as in the present investigation, the *S*s are presented with the judgment or opinion of another person which is to varying degrees different from the *S*'s judgment. In employing this operation, no additional statements are made directing, recommending, suggesting or in any other way indicating the desirability of change on the part of the *S*s. Although advocacy of change may be implicit in the presentation of the other person's response, only empirical evidence can demonstrate that this is so. A forthcoming paper will consider the possibility that different processes are involved in change produced by opinion discrepancy and by advocacy of change.

[3] Fisher and Lubin (1958) present data indicating that with extreme amounts of opinion discrepancy little change is obtained.

RESULTS

After Bartlett's test indicated no basis for rejecting the assumption of homogeneity of variances, an analysis of variance was performed on the change scores. Table 1 summarizes this analysis and reveals that both main effects and the interaction are significant at acceptable levels.

The significant mean square for interaction suggests that the relative effects of opinion discrepancy are in some way dependent on external commitment. Statistical tests indicate no reliable differential effects of opinion discrepancy under conditions of either congruent or incongruent commitment, although the data in Table 2 suggest a slight increase in actual change with increasing magnitude of opinion discrepancy. Under the control condition, however, the mean change (1.56) for for the 3-unit opinion discrepancy condition is significantly larger ($p < .01$) than the mean changes (.63 and .73) for the 1-unit and 2-unit opinion discrepancy conditions.

In spite of the reliable interaction effect, the magnitude of the F ratio (27.94) for commitment conditions justifies the inference that this factor produces reliable differences in mean change under all three degrees of opinion discrepancy. The average change (.97) under the control condition is significantly different from the average change (.54) under the congruent condition and from the average change (1.49) under the incongruent condition. The direction of these differences supports the contention that congruent external commitment produces, in general, resistance to change and that incongruent commitment tends to facilitate change.

Inspection of overall mean changes under the three conditions of opinion discrepancy reveals the same pattern found by Hovland and Pritzker (1957), i.e., the greater the opinion discrepancy the greater the actual change.

TABLE 1
ANALYSIS OF VARIANCE OF OBTAINED CHANGE SCORES

Source	*df*	Mean Square	*F*	*P*
External commitment	2	784.04	27.94	<.001
Opinion discrepancy	2	216.17	7.70	<.01
Interaction	4	71.46	2.55	<.05
Within	63	28.06		

TABLE 2
MEAN OBTAINED CHANGE

External Commitment	Opinion Discrepancy Units			
	1	2	3	Combined
Congruent	.45	.54	.63	.54
Control	.63	.73	1.56	.97
Incongruent	1.26	1.58	1.63	1.49
Combined	.77	.95	1.27	

As previously noted, however, this pattern is associated primarily with the commitment control condition and even there it is reliably established only between the conditions of 2-unit and 3-unit discrepancy. This last result was obtained with a considerably smaller *N* than Hovland and Pritzker employed.

Evaluation of amount of change obtained in terms of the initial positions held by the *S*s yielded essentially the same pattern of results for those holding extreme, moderate, or more neutral initial positions. Similar results were also obtained when differences in directionality of initial opinions were examined.[4] Previous research (Hovland, Janis, & Kelley, 1953) has indicated that resistance to change is frequently manifested by contracommunication or "boomerang" responses. Table 3 presents the frequency with which individuals made one or more boomerang responses in each combination of conditions. Chi square analysis indicates a nonsignificant effect of the various conditions employed on boomerang responses.

Perhaps, it might be suggested, the final responses of the *S*s can be interpreted without

---

[4] Statistical analyses of these results are precluded because of extreme disproportionality of cell entries and absence of a control group to account for possible regression.

TABLE 3

FREQUENCY OF APPEARANCE OF ONE OR MORE BOOMERANG RESPONSES PER *S*

External Commitment	Opinion Discrepancy Units		
	1	2	3
Congruent	7	7	4
Control	4	6	4
Incongruent	5	3	3

TABLE 4

ANALYSIS OF VARIANCE OF DEVIATION SCORES

Source	*df*	Mean Square	*F*	*P*
Congruence	1	105.13	10.67	<.01
Separation of check marks	1	94.53	9.60	<.01
Interaction	1	4.50	—	—
Within	28	9.85		

reference to initial positions; that is, if the final responses are related to attribution and opinion discrepancy, statements concerning external commitment may be gratuitous. Certain conditions within the present experiment offer data that make possible an examination of this suggestion. Under the condition of congruent commitment and one-unit discrepancy and that of incongruent commitment and two-unit discrepancy, *S*s were presented with patterns of check marks that were essentially identical. The same similarity obtained between the conditions of congruent commitment and two-unit discrepancy and of incongruent commitment and one-unit discrepancy. Along with the variation of congruence these two sets of conditions differed in the amount of separation of the check marks representing attribution and opinion discrepancy. In the former conditions one unit separated the two check marks, while in the latter conditions a two-unit separation was presented.

If final response is related only to the pattern of check marks no difference should appear between conditions that are differentiated only by the labels, congruent and incongruent. If differences between congruent and incongruent commitment conditions do appear, reference to initial positions is appropriate.

In order to test these alternatives the four conditions indicated were analyzed in a 2×2 factorial design in which one independent variable can be described in terms of congruence and the other in terms of separation of check marks. The dependent measure was the average absolute deviation for each *S* of final responses from the midpoint between the two check marks representing the attribution and the opinion discrepancy. Table 4 presents the summary of the analysis of variance of the deviation scores.

The results indicate that congruence is related to final responses. Within the limits of the present data, it appears that initial position is appropriately considered in evaluation of final responses and that statements concerning external commitment are not gratuitous. Final response is also related to the amount of separation between the two check marks, which is in part reflective of the effect of opinion discrepancy.

DISCUSSION

The results of the present experiment taken together with those of the earlier study (Rosenbaum & Zimmerman, 1959) support the contention that the attribution of an opinion that is congruent with a person's current response tendency produces resistance to a subsequent attempt to influence his opinion. These results are comparable to those obtained under self-commitment procedures, and therefore the process would seem to be appropriately named *external* commitment.

In accounting for the positive results of the earlier study, use was made of the interpretation employed by Hovland et al. (1957) as to the basis for commitment effects. Consistency of behavior is viewed as a desirable characteristic for which socially mediated rewards are dispensed. Under self-commitment procedures, the person himself states the standard to which he is to remain consistent. Failure to do so is frequently socially punished by such terms as "inconsistency," "wishy-washy," "mixed-up," "neurotic," etc. In situations involving external commitment the standard is set by the source of the attribution and can be regarded as a cue to which a matched response has in the past led to social reward for most persons in our culture. In effect, the external source is stating a standard for a particular response on the part of the recipient of the attribution. However, in the previous study (Rosenbaum & Zimmerman, 1959) and in the current experiment, conflict arises from the fact that not only does the external source present a standard in the form of the attribution, but he provides a second standard in presenting his own position. Potential approval would seem to be offered for presenting a response similar to that of the external source as well as for responding in accord with the attribution. Whether the attribution standard or the discrepancy standard governs the final response presumably depends on undetermined factors associated with the person's response hierarchy. For an unselected group the mean change should appear as a compromise response falling between the two standards. An increase in obtained mean change should then occur, varying directly with the amount of opinion discrepancy. Although a tendency toward a positive relationship is indicated by the results for different amount of opinion discrepancy in the congruent commitment conditions, the relationship is clearly not significant, in spite of evidence in the no commitment condition that presenting the response of the external source is effective. It would then appear that, at least in the present situation, the variable of congruence of commitment is considerably stronger in its effect than that of opinion discrepancy.

In the situation of incongruent commitment *S* faces a somewhat different set of circumstances in that the two standards set by the external source are in considerably greater alignment than in the congruent condition. Both the standard represented by the response of the other person and the commitment standard are presented in a direction away from the *S*'s original position, although they are completely aligned only in the case of three-unit opinion discrepancy. Again assuming relative equality of response potential among persons for each standard, the mean amount of change obtained can be expected to increase as the amount of opinion discrepancy increases. Here, too, the results indicate an insignificant tendency toward a positive relationship. Incongruent attribution seems also to affect response more strongly than does opinion discrepancy.

Why the effect of incongruent commitment in the present experiment differs from that in the earlier one is not clear. It would appear that the amount of opinion discrepancy alone did not produce the limiting effect on incongruent commitment found in the earlier study, although it may be necessary to explore the effects of more extreme amounts of opinion discrepancy before a conclusive statement becomes possible.[5]

It was suggested that the amount of emotionality or ego involvement aroused by opinion topics may be a critical variable affecting response under incongruent commitment. This variable was not directly treated in the present study although items less likely to arouse such response than those of the earlier study were used. It remains possible that some interrelationship between this aspect of the opinion topics and the amount of opinion discrepancy produced the constriction on the effect of the incongruent commitment treatment in the earlier study.

An alternative context for interpreting the results pertaining to attribution of position may be available. Presenting situational cues that make a person's membership in a particular group salient favors responses similar to those learned and adopted in the context of that membership. Charters and Newcomb (1952) demonstrated that Catholics were more likely to express opinions reflective of orthodox Catholicism when such cues were present in an otherwise nondenominational context than when membership cues were absent. Kelley (1955) offered similar evidence indicating greater resistance to an attempt to change opinions of Catholic high school students when membership cues are present. In the present experiment, attribution of position can be regarded as having increased the salience of *S*'s previous opinion responses.

This interpretation is more immediately plausible in the case of congruent attribution than for incongruent attribution. In the latter instance, a response hierarchy may be assumed that includes several alternative responses in addition to the pre-experimental questionnaire response, among them one similar to that phrased in the incongruent attribution. The situational presentation of the attribution may then tend to elicit the response corresponding to this previously nonapparent portion of the response hierarchy. In group terms, it is as if one of a person's multiple membership groups is made salient rather than another. From a general S-R interpretation, the response that is obtained is the one associated most strongly with the cues in the particular situation, which is in turn dependent on the person's learning history.

---

[5] An unpublished pilot study similar to the present one in most respects but presenting one unit, three units, and five units of opinion discrepancy to *S*s holding extreme positions yielded similar findings to those reported above.

SUMMARY

The present study was concerned with the effect of prior attribution of opinion to a person on a subsequent attempt to influence his opinion. It was hypothesized that an attribution congruent with the recipient's current opinion results in resistance to influence while an incongruent attribution facilitates opinion change. The effects were predicted on the basis of a process termed "external commitment."

An earlier study had suggested that limits on the effects of external commitment may depend on the amount of opinion discrepancy involved in the influence attempt. In the experiment reported, three degrees of opinion discrepancy were systematically presented with three external commitment treatments; congruent attribution, incongruent attribution, and a control or no commitment condition.

The results support the hypothesized effects of external commitment. Within the situation studied, external commitment has a stronger effect than opinion discrepancy, although results for a control condition offer partial support for Hovland and Pritzker's (1957) finding that the greater the opinion discrepancy, the greater is the amount of change obtained.

The process of external commitment was discussed as involving socially mediated rewards for consistent behavior. An alternative interpretation treated attribution of position in terms of salience of an earlier response tendency.

REFERENCES

CHARTERS, W. W., & NEWCOMB, T. M. Some attitudinal effects of experimentally increased salience of a membership group. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (Rev. ed.) New York: Holt, 1952. Pp. 415–420.

DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.

FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *J. abnorm. soc. Psychol.*, 1958, **56**, 230–238.

FISHER, S., RUBINSTEIN, I., & FREEMAN, R. W. Intertrial effects of immediate self-committal in a continuous social influence situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 200–207.

GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.

HOVLAND, C. I., CAMPBELL, E. H., & BROCK, T. The effects of "commitment" on opinion change following communication. In C. I. Hovland (Ed.), *The order of presentation in persuasion.* New Haven: Yale Univer. Press, 1957.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.

HOVLAND, C. I. & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.

KELLEY, H. H. Salience of membership and resistance to change of group-anchored attitudes. *Hum. Relat.*, 1955, **8**, 275–290.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity and the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

POWELL, N. J. *Anatomy of public opinion.* New York: Prentice-Hall, 1951.

ROSENBAUM, M. E., & ZIMMERMAN, ISABEL M. The effect of external commitment on response to an attempt to change opinions. *Publ. opin. Quart.*, 1959, **23**, 247–254.

YEAGER, W. H. *Effective speaking for every occasion.* (2nd ed.) New York: Prentice-Hall, 1951.

(Received May 6, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 21-24

A CONSTANT ERROR APPROACH TO THE STUDY OF DIMENSIONS OF SOCIAL PERCEPTION[1]

LEON H. LEVY
Indiana University

AND ROBERT D. DUGAN
State Farm Insurance Companies

RATINGS may be treated either as descriptive data to be used in the further study of the people rated, or they may be treated as behavior in their own right to be studied for what they may reveal about those performing the ratings and the processes involved in this task. It is with this latter use of ratings that this study is concerned.

The observation that ratings of individuals on a number of different traits tend to correlate with each other to a degree that seems unwarranted has given rise over the years to such concepts as the "halo effect" and the "logical error." Thorndike (1920), in coining the term "halo effect," referred to it as a constant error in psychological rating resulting from "suffusing ratings of special features with a halo belonging to the individual as a whole." Guilford (1954) described the logical error, earlier observed by Newcomb (1931), as an instance of judges giving "similar ratings for traits that seem logically related in the minds of raters."

In each of these cases, the same individuals are rated on a number of different traits and the observed intercorrelations between these traits has been assumed to reflect a constant error. However, given the procedure and assuming that the intercorrelations do not represent the true state of affairs between the traits involved, it is impossible to separate the operation of the halo effect and the logical error from each other, or from that of yet another equally tenable explanation. This is that the correlations may arise because the traits rated are not independent of each other from the standpoint of the perceptual process involved in making the judgments, but rather represent specific instances of a more limited number of dimensions of judgment or social perception.

This latter consideration would suggest that from an analysis of the nature of the constant errors or intercorrelations found in trait ratings, information might be obtained concerning the nature of certain aspects of social perception. This approach has been followed, in effect, by Asch (1946) and many after him. However, in none of these studies has the investigator controlled for the operation of the so-called logical error. In fact, one might consider most of them as concerned with exploring the nature of what Bruner and Tagiuri (1954) so aptly refer to as the "naive, implicit 'theories' of personality people work with when they form impressions of others."

Whereas in previous studies demonstrating correlations between trait ratings, these relationships could be said to reflect the experience of the rater such that he has learned that Trait A implies Trait B (Hays, 1958), the present study is concerned with the question of whether certain relationships exist between trait ratings which *cannot* be attributed to such learning, but must rather reflect more basic dimensions of social perception. The entire conceptual structure of the projective approach to the understanding of perception and personality rests upon the assumption that perception is functionally related to certain states of the perceiver and, hence, would suggest the existence of such dimensions as perceptual counterparts of these states. This study represents a first attempt at determining their nature.

METHOD

One hundred introductory psychology students (52 men and 48 women) served as subjects (*Ss*) in the experiment. Each was seen individually and required to make a dichotomous trait judgment in connection with a particular bipolar trait for each of 225 photo-

---

[1] This investigation was supported in part by a research grant (M-2067) from the National Institute of Mental Health, of the National Institutes of Health, Public Service.

The authors wish to thank Jane A. Kelly for running the subjects of this study.

g-427
8.9.70

TABLE 1

INTERRELATIONS BETWEEN TRAITS

Traits	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Good–bad	50	48	41	49	44	42	50	33	29	13	29	00	07	−02
2 Safe–dangerous		40	34	42	29	33	18	22	13	12	06	−02	13	−01
3 Kind–cruel			23	23	26	32	16	31	22	22	04	10	11	−05
4 Reliable–unreliable				55	49	45	40	30	36	16	14	03	14	01
5 Honest–dishonest					45	42	37	38	13	12	06	−12	16	14
6 Stable–unstable						34	35	25	17	25	31	04	19	03
7 Friendly–unfriendly							31	36	19	21	05	07	18	−07
8 Ambitious–lazy								09	11	19	16	−13	06	−03
9 Generous–tight-fisted									19	24	23	15	18	04
10 Happy–unhappy										04	10	26	18	17
11 Thoughtful–self-centered											12	20	17	07
12 Attractive–unattractive												14	08	−06
13 Relaxed–tense													−13	00
14 Warm–cold														13
15 Unanxious–anxious														

Note.—An *r* of .19 is required for significance at the .05 level and an *r* of .25 for significance at the .01 level.

graphs presented to them one at a time. *S*s were instructed to make their judgments on the basis of their first impression of the person. The photographs were of white male applicants for admission to graduate training in psychology at Indiana University who, for one reason or another, never became students in the program. Each photograph was mounted on a 3 × 4 in. piece of poster board and was assigned at random to one of 15 decks. All photographs within a deck were judged on the same trait and had the two poles of this trait typed on a label attached below the picture. For all *S*s the same picture was always judged in connection with the same trait, but the order of presentation of pictures varied randomly from *S* to *S*. Thus, each *S* judged all 225 photographs but made only one judgment for each.

The 15 traits rated may be seen in Table 1. Traits 2, 11, 13 and 15 were typed in the reverse order from that presented in the table when they were presented along with their accompanying pictures to *S*s.

Each *S* was given a score for each trait which represented the number of photographs judged in the positive direction. Therefore scores for each trait could range from 0 through 15.

The method of the present study permits a clear distinction to be made between the operation of logical error and any of its kindred concepts, and perceptual dimensions. Since logical error results from inferring a certain amount of one trait from a certain amount of another trait in the same person, it cannot be invoked to account for correlations between traits obtained through the method described here where the two traits were rated not as characteristics of the same person, but rather as characteristics of different persons. For having judges rate one variable per photograph means that the judges cannot consider one variable or trait about a photograph to imply another variable or trait about another photograph. By the same reasoning, such correlations could not be assumed to represent the true relationship between the traits involved for the persons rated. Hence, any correlations found between traits in the present study can only be attributed to certain aspects of the judges or the perceptual processes involved, viz., the existence of certain fundamental dimensions of perception which are variously represented by the traits being judged.

RESULTS AND DISCUSSION

Pearson product-moment correlations were calculated between the 15 obtained scores and the resulting matrix was subjected to a factor analysis in which the Thurstone Multiple Group Method of extraction was used. The resulting factors were rotated graphically to orthogonal simple structure with communality estimates corrected by the Wherry Iterative Method (Wherry, 1949). The graphic rotation was performed using only the factor loadings and a visual criterion of simple structure rather than using the variable names and meanings. The table of intercorrelations and the rotated factor loadings are presented in Tables 1 and 2.

From the intercorrelation matrix it is obvious that relationships do exist between attribution tendencies for certain traits. If these ratings had all been performed on the same persons it would have been reasonable to account for many of the correlations in terms of synonymity or logical implication or true relationship between the two traits involved. In this instance these explanations are not tenable, and so it is reasonable to propose as an explanation the existence of certain dimensions of perception which are responsible for the correlations obtained.

Although we recognize the likelihood that

TABLE 2
ROTATED FACTOR LOADINGS

Traits	Factors				
	I	II	III	IV	h^2
1 Good-bad	.75	.32	.07	.06	.67
2 Safe-dangerous	.56	.37	.02	.05	.45
3 Kind-cruel	.45	.48	.07	.08	.44
4 Reliable-unreliable	.65	−.09	.36	.00	.56
5 Honest-dishonest	.65	.03	.30	−.15	.54
6 Stable-unstable	.62	−.02	.30	.14	.49
7 Friendly-unfriendly	.56	.12	.17	.05	.36
8 Ambitious-lazy	.49	.02	.17	.00	.27
9 Generous-tight-fisted	.44	.12	.11	.18	.25
10 Happy-unhappy	.33	.08	.08	.23	.17
11 Thoughtful-self-centered	.27	.04	.07	.34	.20
12 Attractive-unattractive	.25	.00	.05	.36	.19
13 Relaxed-tense	.00	.07	−.05	.49	.25
14 Warm-cold	.21	−.03	.10	−.04	.06
15 Unanxious-anxious	.02	.12	−.02	.01	.02

the factorial structure is dependent upon the limited number of traits sampled, we propose that the factors extracted from the matrix of intercorrelations are representative of certain major dimensions of social perception. From this standpoint, Factor I, which is a general factor, has been labeled General Evaluative Bias. This factor would seem to suggest that there is always a "halo effect" operating in that individuals differ in their tendency to react positively or negatively to another person and that this tendency is manifest throughout the realm of those traits sampled which permit an evaluative judgment to be made.

Factor II has been labeled Harmfulness, suggesting the importance of potential threat or hazard in the perception of others. Factor III seems most appropriately designated as Dependability in that a person described as reliable, honest, and stable, would be one that could be depended upon whereas the contrary would be true for a person described in opposite terms. Factor IV is somewhat more difficult to define, but has been tentatively labeled Affability.

We propose that these four factors, while not in any way exhausting the realm of possible dimensions of social perception, may be taken as representing four questions that each person asks about another person, and that there are individual differences in their tendencies to answer them in the affirmative or negative. These questions might be paraphrased as follows: for Factor I, "Is he good or bad?"; for Factor II, "Is he likely to cause me any trouble?"; for Factor III, "Can I count on him?"; and for Factor IV, "Would I want him as a friend?"

For the personality theorist the present study might be taken as demonstrating that, for each of the dimensions isolated, persons vary in their adaptation level (Helson, 1955), or point of subjective equality, for their judgments of others. This interpretation immediately raises the question of the personality variables of which this parameter may be a function as well as that of those behaviors which may be expected to vary as a consequence of variations in it. Rather than speculate on the answers to these questions at this point, we simply raise them to indicate the directions for further research suggested by the present study.

In contrast to the findings of Asch (1946) and those of others following him, the present study gives no support for the centrality of the "warm-cold" dimension. Its only significant correlation is with "stable-unstable," and it fails to have a significant loading on any of the factors extracted. Since the distribution of scores on this dimension had a range from 4 to 12, the finding cannot be attributed to restricted variance and must therefore find its explanation elsewhere. In order to determine whether the unreliability of this dimension precluded its correlating with other dimensions, each *S*'s scores for his odd and even numbered pictures were correlated. The resulting correlation, corrected by the Spearman-Brown formula, was .32 suggesting that reliability, while not high, cannot be held entirely accountable for the lack of correlation found. One possible explanation for the discrepancy may reside in the difference in procedures used in this study and in previous ones. It may be that different processes were tapped in each. Thus, in studies on the Asch model *S* was always supplied with information regarding this dimension, such studies could only reflect the possible personality-perceptual bases responsible for relationships between "warm-cold" and other dimensions. In the present study, on the other hand, relationships between dimensions or traits can only occur through relationships in those personality-perceptual processes responsible for differential tendencies in ascribing them to people in a context of minimal information.

The complete absence of correlation between "unanxious-anxious" and any of the other traits is somewhat more puzzling. Again, from inspection of the distribution of scores for this trait, this finding cannot be explained in terms of restricted variance, and the odd-even reliability, corrected by the Spearman-Brown formula, was in this case .37. It may be that some *Ss* interpreted anxiety as eagerness rather than dysphoria, or it may be that if additional traits were included anxiety would emerge as part of another major dimension. Until further research is carried out these questions must remain unanswered.

The immediate practical implications of the present findings reside in the use of rating scales in the description or evaluation of persons. To the operation of the halo effect and the logical error, there must now be added the dimensions of social perception operative when ratings are made and which may also be responsible for some portion of the correlations found between ratings. Two ratings may correlate simply because they represent the same perceptual dimension. This is of particular importance where correlations between trait-ratings form the bases for discussions of personality types and syndromes (e.g., Sheldon & Stevens, 1942; Lorr, Rubinstein, & Jenkins, 1953; Cattell, 1957). Until it has been demonstrated that the judges are capable of producing ratings which are independent of each other, the personality types described by this procedure may be expected to reflect the perceptual processes of the judges as much as they do the nature of the objects of their judgments.

SUMMARY

A method was described for use in the study of dimensions of social perception which consisted of the factor analysis of intercorrelations between trait judgments of photographs where each judgment was made of a different stimulus. The method precludes the possibility that correlations between judgments represent learned contingencies and therefore may be taken as a means of isolating basic social perceptual dimensions that in turn may represent the perceptual counterparts of certain personality parameters of the judges.

Four factors were extracted from an intercorrelation matrix representing judgments on 15 bipolar traits. These were labeled as follows: Factor I—General evaluative bias; Factor II—Harmfulness; Factor III—Dependability; Factor IV—Affability.

The relationship between these findings and those of previous studies in social perception was discussed briefly, as was their implications for research using personality ratings in the description of personality types and syndromes.

REFERENCES

ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258-290.

BRUNER, J., & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. Cambridge, Mass.: Addison-Wesley. 1954. Pp. 634-654.

CATTELL, R. B. *Personality and motivation structure and measurement*. Yonkers, N. Y.: World Book: 1957.

GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.

HAYS, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Pp. 289-299.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1955, **55**, 297-313.

LORR, M., RUBINSTEIN, E., & JENKINS, R. L. A factor analysis of personality ratings of outpatients in psychotherapy. *J. abnorm. soc. Psychol.*, 1953, **48**, 511-514.

NEWCOMB, T. An experiment designed to test the validity of a rating technique. *J. appl. Psychol.*, 1931, **22**, 279-289.

SHELDON, W. H., & STEVENS, S. S. *The varieties of temperament*. New York: Harper, 1942.

THORNDIKE, E. L. A constant error in psychological rating. *J. appl. Psychol.*, 1920, **4**, 25-29.

WHERRY, R. J. A new iterative method for correcting erroneous communality estimates in factor analysis. *Psychometrika*, 1949, **14**, 231-241.

(Received April 17, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 6z, No. 1, 25-30

A COMPARISON OF SEMANTIC STRUCTURES IN AMERICAN SOUTHWEST CULTURE GROUPS[1]

GEORGE J. SUCI[2]

Gerontology Branch, National Heart Institute, and National Institute of Mental Health

An object is perceived, or a concept described, in terms of a set of attributes, and most attributes, but not necessarily all, have linguistic equivalents that permit their communication. This assumption is, in part, the basis for an operational definition of meaning—the semantic differential, extensively described elsewhere (Osgood, Suci, & Tannenbaum; 1957). Briefly, concepts or objects are judged on rating scales defined with adjective opposites (e.g., *good–bad*). In English, adjectives appear to be related such that only a few dimensions are needed to account for most variation in such judgments. These dimensions define a semantic space—and the meaning of a concept is given by the coordinates of its location in this space.

A concept has the same meaning for two individuals or two culture groups when the concept is located in the same position of the respective semantic spaces. That such an equivalence between concepts can be said to exist depends on the equivalence of the semantic spaces; i.e., the dimensions, and therefore attributes, must mean the same for the two individuals. To avoid circularity, the meaning of attributes cannot be inferred from the meanings, or locations, of the concepts; instead, such meaning is inferred from the totality of relations among the attributes themselves. If the same relations exist between a set of attributes for two different individuals, their semantic spaces are equivalent and their meanings of a concept may be compared. Obviously without such equivalence of space, communication between individuals regarding the meanings of their concepts is seriously impeded.

Extremely different subjects have shown highly similar semantic structures. Suci (1952) found that a group of high and a group of low scorers on the California F Scale judged ethnic concepts in similar semantic frames of reference. The meanings of the ethnic concepts, however, differed significantly between the groups. The same kind of results were demonstrated when Republican and Democratic subjects rated politically relevant concepts (Osgood et al., 1957, pp. 104–124). Meanings of concepts like FDR, MCCARTHY, POLICY IN CHINA differed significantly while the attribute relations remained constant. Similar results are reported by Bopp (1955) for normal and schizophrenic subjects.

Does semantic constancy occur across subjects selected from different cultural and linguistic backgrounds? Kumata and Schramm (1956) showed equivalent semantic structures for Korean bilingual exchange students, Japanese bilingual exchange students, and American natives when the semantic judgments were made in the respective languages. Later Kumata (1957) showed the same kind of equivalence for Japanese and American monolinguals, thus extending the generality of the Kumata and Schramm findings. Triandis and Osgood (1958), aiming more directly at a test of Whorf's hypothesis (1956) that language determines cognition, found that Greek and American monolinguals used similar semantic spaces. Thus, Triandis and Osgood, as well as Kumata, have produced evidence against the Whorfian hypothesis, at least with respect to meaning as measured by the semantic differential. The present study extends the comparison of semantic structures to subjects from the American Southwest—Spanish, Hopi, Zuni, and Navaho subjects.

---

[1] This study was conducted as a part of the Southwest Project in Comparative Psycholinguistics, sponsored by the Committee on Linguistics and Psychology of the Social Science Research Council under a grant from the Carnegie Corporation of New York. The data were analyzed at the University of Illinois Computer Laboratories. For a more complete description of the entire project, see Casagrande (1956). The author wishes to express his appreciation to the following members of the project who were kind enough to collect the data used in this study: Susan Ervin for Navaho; J. B. Carroll, Lois Elliott, and H. S. Maclay for Hopi; J. R. Bastian, Sally Bastian, E. H. Lenneberg, and S. S. Newman for Zuni; and S. Saporta for Spanish.

[2] This study was carried out when the author was at the University of Illinois.

Method

In general, the design consisted of constructing a semantic differential and administering it to subjects (*S*s) from each culture group, intercorrelating the semantic scales for each group, factoring each of these sets of relationships, and comparing the obtained factor structures. The Southwest groups will be designated as "experimental" groups to differentiate them from an English-speaking "control" group which is described below.

The Semantic Differential

A first selection of bipolar adjective scales was made as follows. Each of 42 nouns was translated into the native language by an informant from each experimental group. The nouns were read one at a time as stimuli to each informant who was instructed to respond with as many appropriate adjective forms as he could. Each "adjective" obtained in this fashion was read back to the informant who was instructed to respond with a proper opposite. Of the total number of bipolar opposites obtained in this fashion, 25 were common to three of the experimental groups. These were retained as items in the semantic differentials given to the *S*s. In the field it was necessary for each investigator to adjust his list because of poor translations, or *S*s' refusals to respond to some of the items. This resulted in the use of different polar terms by different investigators, and, in the final analysis, only 15 scales were common to all groups.

Ten concepts were selected to be rated on the scales. As with the scales, many of the concepts were found unsatisfactory because of translation difficulties or because they touched on sensitive areas and threatened rapport with the *S*s; therefore, new concepts had to be substituted in some cases and different sets of concepts resulted for some groups. *S*s of all groups rated: MEXICAN, CORN, MAN, WOMAN, HORSE, and RAIN. The Zuni group also rated COYOTE, RED, YELLOW, and BLUE-GREEN. The Hopi and Navaho also rated NAVAHO, WHITE PEOPLE, COYOTE, and SELF, and the Spanish group rated RELIGION, DEATH, ARMY, and GODFATHER.

Control for Concept Differences

Semantic structures may vary with different concepts (Osgood et al., 1957, pp. 176–188); therefore, it was necessary to estimate the extent of semantic space variation due to differences in concepts rated by each group. All the different concepts were presented to a "control" group of English speakers, monolingual college students at the University of Illinois. Three semantic spaces were factored out, each one based on a subset of concepts rated by an experimental group. The three semantic spaces were compared. Since each English speaking *S* was his own control, differences between these structures could be attributed only to differences in concepts, not to differences in language or culture; therefore, factor differences between the control semantic structures implied that comparisons between experimental groups were not valid for those factors. In the analysis, no comparisons between experimental groups were made when such differences occurred.

Subjects

The experimental *S*s were 26 Zuni, 28 Hopi, and 27 Navaho Indians, and 32 Spanish speakers from the Albuquerque, New Mexico area. All *S*s were fluent speakers of their native language, but English was spoken with varying degrees of competence. The Spanish, Hopi, and Zuni *S*s were bilingual but dominant in their own language as judged by the linguists working with these groups. The extent of bilingualism in the Navaho *S*s was judged on the basis of a bilingualism test.[3] Of the 27 Navaho, 14 were English dominants, 6 were Navaho dominants, and 7 were Navaho monolinguals. The median age and age-range for each group was as follows: Zuni, 35, 12–70; Hopi, 24, 12–64; Navaho, 45, 17–75; Spanish, 26, 18–34; English speakers, 21, 19–26. The sex distribution, male and female respectively, in each group was as follows: Zuni, 15, 11; Hopi, 12, 16; Navaho, 8, 19; Spanish, 21, 11; American, 18, 15.

Instructions and Procedure

The literate Spanish and control groups received standard instructions (Osgood et al., 1957, pp. 82–84). Administration was in a group using paper-and-pencil forms printed in the respective native languages.

Since the Indian people are nonliterate in their native languages, *S*s from these groups could not respond to the standard rating procedure and a modified technique had to be devised. Twenty-five lines, one for each scale, with the ends and the middle of each line marked, were mimeographed on legal size paper. The lines were six inches long. The instructions, given below, will explain the procedure employed in the administration of the items to the Indian *S*s.

> We would like to know how you describe or talk about certain things. For example, you may think of something being very fast or very slow, or a little bit strong or a little bit weak. Suppose we wanted

---

[3] The bilingualism test was developed by Susan Ervin.

TABLE 1

FIRST THREE FACTORS AFTER BIQUARTIMIN ROTATION

Scales	Navaho			Hopi			Zuni			Spanish			A-N-H[a]			A-Z[b]			A-S[c]		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
heavy-light	07	47	−01	09	45	−29	11	10	53	−35	34	07	−04	47	33	01	12	02	−21	49	−12
hard-soft	−26	26	−05	−01	30	−25	−08	23	−08	−28	22	26	−29	28	−10	−38	45	−05	−07	30	−18
pretty-ugly	51	07	−02	72	01	−14	64	−06	00	79	00	−01	61	−06	10	66	01	06	41	−03	27
hot-cold	−40	40	08	20	−01	−39	20	−16	42	42	14	38	01	00	54	−05	02	59	−08	05	65
clean-dirty	34	26	00	58	−07	14	65	03	10	61	08	−12	75	−05	−12	72	03	−12	67	−06	−10
dry-wet	−12	−15	28	−05	−22	−43	12	−18	−15	−13	−11	23	−20	01	15	−29	22	37	−09	−01	17
industrious-lazy	41	01	24	48	05	07	16	67	−02	15	62	−06	34	32	−04	17	66	07	29	46	17
happy-sad	24	42	−13	53	07	02	59	08	−02	53	11	08	39	−06	31	37	00	46	19	03	53
rich-poor	00	49	−06	54	03	07	60	11	15	33	11	−43	61	24	−15	59	26	−07	61	16	−18
sweet-sour	32	02	02	21	−08	47	33	24	17	78	−03	06	58	−20	15	63	−16	10	44	−22	29
long-short	05	48	03	13	33	08	23	31	21	−04	51	−01	13	54	−03	14	36	24	07	43	−30
fast-slow	03	−03	65	−02	53	01	−10	71	12	08	45	−10	08	60	−10	−03	63	−07	09	48	08
strong-weak	12	46	−07	01	64	05	02	72	−04	−03	64	00	01	58	16	−04	54	11	02	64	10
straight-crooked	05	05	00	06	−19	05	50	05	−05	38	20	−05	53	17	21	55	12	09	48	08	04
good-bad	53	18	01	42	05	34	31	46	−26	77	02	−01	69	02	12	71	−02	00	52	06	25
% Common Variance	14.9	17.3	7.1	25.1	15.4	11.4	28.8	26.4	8.9	41.6	19.9	6.4	35.9	19.6	8.7	36.1	20.3	10.1	22.1	17.2	13.8

Note.—Decimal points are omitted.

[a] American control for Navaho and Hopi concepts.

[b] American control for Zuni concepts.

[c] American control for Spanish concepts.

to know how strong or weak something like MOUNTAIN is. You could think of this end (experimenter [*E*] points to one end of a line) as being something very, very strong and this end as something very, very weak. Now since MOUNTAIN is very, very much like the very, very strong, it would be very much like the very, very strong thing on this end and you would show me, by pointing (or pencil mark), that this is where MOUNTAIN belongs.

On the other hand, something like GRASS is not strong, but quite weak, and you would show me this by placing it close to this thing which is very, very weak. A TREE is quite strong but certainly not as strong as MOUNTAIN, so you might put TREE here, on the strong side, but not as close to the very, very strong end as MOUNTAIN. A BUSH is not as strong as a TREE, but neither as weak as GRASS, you may put it between GRASS and the TREE, here, but a little on the strong side. (The *E* repeats the above with BREAD, BLANKET, SPIDER, and SNAKE, in that order form good to bad, on scale *good–bad*.)

Now think of this end as very, very active and this other end as very, very passive. Now where do you think MOUNTAIN should go? COW? DEER? BEE?

Now here are a number of lines. I will tell you for each line what each end is. Then I will ask you to put something like RAIN where you think it belongs on the line. There is only one place on the line where you cannot put the thing you are to describe, and that is on this point (*E* points to center). You may put it close to this point, but not on it (this is to help force the *S* to make a discrimination).

The above instructions were followed as closely as possible by each investigator. One concept was presented with each sheet of scales. The ends of the scale and the concept were read to the *S* for each judgment. The items were given in the native language to all *S*s. English instructions were used except with the Navaho who were instructed in Navaho.

RESULTS

Digits 1 through 7 were assigned in left to right order to the seven points on the scales of the English and Spanish semantic differentials. A score was the digit corresponding to the check-mark representing the *S*'s rating. The Indian differentials were scored by measuring the distance, in one-tenth inches, from the left side of the scale to the check-mark locating the concept. Each *S*'s rating formed a scale by concept matrix of order 15 × 10. Since

TABLE 2

COEFFICIENTS OF FACTORIAL SIMILARITY

Factor 1								Factor 2						
N	H	Z	S	N-H[a]	Z[b]	S[c]		N	H	Z	S	N-H[a]	Z[b]	S[c]
1.00	.72	.61	.69	.79	.79	.72	Navaho	1.00	.60	.42	.65	.60	.45	.63
	1.00	.91	.81	.87	.83	.82	Hopi		1.00	.74	.79	.87	.70	.88
		1.00	.83	.91	.88	.86	Zuni			1.00	.84	.76	.80	.79
			1.00	.90	.90	.86	Spanish				1.00	.92	.89	.95
				1.00	.99	.96	American-Navaho-Hopi					1.00	.88	.97
					1.00	.94	American-Zuni						1.00	.90
						1.00	American-Spanish							1.00

Factor 3								Factor 4						
N	H	Z	S	N-H[a]	Z[b]	S[c]		N	H	Z	S	N-H[a]	Z[b]	S[c]
1.00	−.15	.11	.03	−.05	.08	.15	Navaho	1.00	.64	.02	−.23	−.15	.18	−.13
	1.00	−.27	−.55	−.32	−.42	−.09	Hopi		1.00	−.24	−.32	−.50	.25	−.28
		1.00	.12	.45	.17	.08	Zuni			1.00	.71	.59	−.17	.66
			1.00	.61	.60	.47	Spanish				1.00	.60	.09	.78
				1.00	.80	.73	American-Navaho-Hopi					1.00	−.16	.86
					1.00	.88	American-Zuni						1.00	−.04
						1.00	American-Spanish							1.00

[a] American control for Navaho and Hopi concepts.

[b] American control for Zuni concepts.

[c] American control for Spanish concepts.

there were N such matrices, a $15 \times 10N$ matrix was formed for each group.

Product-moment correlation coefficients were calculated between scales for each group by summing the cross-products over both *Ss* and concepts so that the number of pairs entering into each coefficient was $10N$. The 15×15 correlation matrix for each group was factored using the principal-axis technique with a communality estimate in the main diagonal, and extracting all 15 factors.[4] The first four principal components were rotated into simple structure using a modification of Carroll's method (1953).[5] Finally, to measure the similarity between the factors, indices of similarity (Wrigley & Neuhaus, 1955) were calculated between related factors.

As shown in Table 2, the first three factors for the control groups were highly similar, but the fourth factors were quite dissimilar. Such dissimilarity implied that differences on Factor 4 could be attributed to concept differences, and that inferences regarding language or culture differences on this meaning dimension could not be made.

The coefficients of similarity for Factors 1 and 2 were consistently positive for all the groups. Factor 3, however, produced some extremely low, and even negative, similarity coefficients. Triandis and Osgood (1958) accepted a value of .75 as the lower limit for statistically significant factorial similarity. Not all coefficients exceeded this value for Factors 1 and 2. The Navaho group consistently produced coefficients less than .75. The other groups, with the exception of Zuni and Hopi whose coefficient for Factor 2 had a border line value of .74, were above this value for Factors 1 and

---

[4] The tables of intercorrelations and the first four principal components for each group have been deposited with the American Documentation Institute. Order Document No. 6238, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] The rotations were performed by J. B. Carroll using an IBM Type 650 program developed by J. B. Carroll, F. Weinfeld, and A. Beaton.

2. None of the experimental groups showed relations above .75 for Factor 3.

From the loadings in Table 1, the first factor was readily interpretable as evaluative with scales like *pretty–ugly*, *clean–dirty* (with the exception of Navaho) and *good–bad* (with the exception of Zuni) producing some of the higher loadings for all the groups. This factor explained the most variance for every group except the Navaho where it accounted for only 14.9%, slightly less than the 17.3% shown by the Navaho second factor. The second factor appeared as a combination of potency (*strong–weak*) and activity (*fast–slow*) for all groups except Navaho who displayed mostly potency scales on this factor with *fast–slow* appearing independently on the third factor. This combination of activity and potency has been termed "dynamism" (Osgood et al., 1957, p. 74) and commonly appears with English speakers. The third factor was not readily interpretable for any group.

DISCUSSION

With the exception of the Navaho group, a high degree of similarity in the semantic structures of *S*s from different cultural backgrounds was found. Since most *S*s were bilinguals, the results do not have the generality of those from the Triandis and Osgood (1958) or the Kumata (1957) studies based on monolingual subjects.

Conclusions are further limited in that only a small amount of the judgmental variance is explained in some cases, especially the Navaho. The first three Navaho factors accounted for only 39% of the reliable variance. In contrast, the three Spanish factors accounted for 68% which concurs with the 70% accounted for by the Greek and English factors (Triandis & Osgood, 1958). In other words, a vast amount of reliable attribute space remains undefined for the Navaho group. To a lesser extent this is also true for the Hopi, whose first three factors account for about 52%. Thus, for the Navaho especially, only a narrow base for comparison with other groups was available. Only further research can tell if this lack of accountable variance explains the Navaho deviation from the other groups. Such limitations, although important and suggestive for further research, should not be permitted to hide the significance of the high degree of similarity. Despite differences in culture, concepts, method of administration and education, similarity between groups was high. Further, the main two dimensions were interpretable as evaluative, and activity-potency, or dynamism, the same factors found with all the other culture groups studied. Thus, taking into account other studies in this series, a high degree of similarity may be assumed to exist in the semantic frames of reference used by subjects of different cultural background.

SUMMARY

The present study is one of a series investigating the similarity of semantic structures for different groups of subjects—in this case *S*s from different American Southwest cultures. The factor structures of a sample of semantic scales indicated that Zuni, Hopi, Spanish, and English-speaking *S*s define a semantic space with similar evaluative and dynamism dimensions. Although the semantic space for a group of Navaho could be defined with the same dimensions, the similarity measures were consistently lower for this group. A third factor for each group could not be interpreted and was highly dissimilar for all groups. Since most *S*s were bilinguals, the generality of the findings was limited in comparison to findings from studies using only monolinguals. Only a small amount of variance is accounted for by the first three factors for some groups, especially the Navaho, suggesting that a large amount of attribute space needs to be defined with further research.

REFERENCES

BOPP, JOAN. A quantitative semantic analysis of word association in schizophrenia. Unpublished doctoral dissertation, Univer. Illinois, 1955.

CARROLL, J. B. An analytic solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.

CASAGRANDE, J. B. The Southwest project in comparative psycholinguistics: A progress report. Soc. Sci. Res. Council, *Items*, 1956, **10**, 41–45.

KUMATA, H. A factor analytic investigation of the generality of semantic structure across two selected cultures. Unpublished doctoral dissertation, Univer. Illinois, 1957.

KUMATA, H., & SCHRAMM, W. A pilot study of cross-cultural methodology. *Publ. opin. Quart.*, 1956, **20**, 229–237.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

SUCI, G. J. A multidimensional analysis of social attitudes with special reference to ethnocentrism. Unpublished doctoral dissertation, Univer. Illinois, 1952.

TRIANDIS, H. C., & OSGOOD, C. E. A comparative factorial analysis of semantic structures in monolingual Greek and American college students. *J. abnorm. soc. Psychol.*, 1958, **57**, 187–196.

WHORF, B. L. *Language, thought and reality: Selected writings of Benjamin Lee Whorf.* New York: Wiley, 1956.

WRIGLEY, C. F., & NEUHAUS, J. O. The matching of two sets of factors. *Amer. Psychologist* 1955, **10**, 418–419.

(Received May 20, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 31-38

PATTERNS OF ROLE DOMINANCE AND CONFLICT IN PARENTS OF SCHIZOPHRENIC PATIENTS[1]

AMERIGO FARINA
Duke University

A GREAT deal of effort has been directed toward determining the possible role played by familial factors in the development of schizophrenia. In past years the role of the mother of the schizophrenic patient has frequently been given primacy (Rosen, 1953; Fromm-Reichman, 1948). However, the nature of the total family structure and the pattern of parental relationships have also received attention, especially in more recent reports. According to these studies, a characteristic interaction pattern of parents of schizophrenic patients is that of greater dominance of one over the other than would be expected in a population of parents of nonschizophrenic offspring (Arieti, 1955; Ellison & Hamilton, 1949; Gerard & Siegel, 1950; Kohn & Clausen, 1958; Lidz, Cornelison, Fleck, & Terry, 1957a; Reichard & Tillman, 1950). Also some preliminary studies have indicated a systematic relationship between the sex of the dominant parent and important patterns of adjustment on the part of the son. In several instances (Rodnick & Garmezy, 1957), it has been found that patients from father-dominated families tend to be married, have numerous friends of both sexes, and generally have made a good premorbid adjustment. Patients from mother-dominated families, on the other hand, tend to be single, social isolates, and to have made a marginal social-sexual adjustment prior to psychosis. These trends are supported by Opler's (1957) finding that patients from mother-dominated subcultural groups tend to be pallidly asexual, whereas those from father-dominated subcultures tend to be more active sexually. Parents of schizophrenic patients, in addition, are reported to relate to each other in a manner characterized by heightened conflict and hostility (Frazee, 1953; Friedlander, 1945; Gerard & Siegel, 1950; Lidz, Cornelison, Fleck, & Terry, 1957b; Tietze, 1949).

This study was primarily designed to investigate the reported association between the sex of the dominant parent and the premorbid adjustment of the schizophrenic son. In addition, the reported conflict in the interaction of these parents as well as the inequality of the dominance relationship seemed to merit investigation. Considerable difficulty is encountered, however, in determining what is usually meant by dominance and conflict in this area of investigation. Many of the studies cited are primarily clinical-qualitative reports and little or no attempt has been made to define these two variables. None of these studies had attempted to measure systematically any aspect of actual parental interaction. In spite of this difficulty, the hope in the present study was to establish operationally defined indices of dominance and conflict which would do minimal violence to commonly held connotations of these terms. Thus, in a standardized situation where the parents are interacting, the more dominant partner can be defined as the one who speaks the most, whose decisions are accepted by the other, etc. In the same situation, extent of conflict can be measured by number of interruptions by each parent, failure to reach agreement, etc. Observing the actual interactions between the parents should have a decided advantage over drawing inferences from paper-and-pencil scales completed by each. If, in addition, the standardized situation were so structured that the parents were unaware of what was being measured, their understandable defensiveness about revealing patterns of conflict and dominance in their relationship could be minimized.

Other studies which have used comparable measures of dominance and conflict give some indication of their usefulness. Matarazzo,

---

[1] This report is based on a PhD thesis submitted to the Graduate School of Arts and Sciences at Duke University. The research was in part financed by a research grant (M-629) from the National Institute of Mental Health. Grateful appreciation is expressed to Norman Garmezy, Eliot H. Rodnick, and Edward E. Jones for their help and suggestions. Thanks are due to the staffs of the State Hospital of Butner, North Carolina and Oteen VA Hospital, Oteen, North Carolina for their help in obtaining *S*s.

TABLE 1

CHARACTERISTICS OF PARENT GROUPS

	Controls	Goods	Poors
Age of Father			
Mean	59.7	55.8	60.5
SD	6.9	9.1	10.2
Age of Mother			
Mean	56.6	52.3	56.7
SD	8.1	8.2	9.7
Years Married			
Mean	37.4	32.0	34.8
SD	7.2	7.0	9.5
Grades Completed by Mother			
Mean	6.4	8.1	8.6
SD	2.8	2.6	3.9
Grades Completed by Father			
Mean	7.3	7.2	9.5
SD	3.0	3.3	3.7
Number of Children			
Mean	5.2	4.2	4.3
SD	2.6	1.8	2.5
Number of Male Children			
Mean	3.4	2.6	2.2
SD	1.4	1.2	1.1

Saslow, and Guze (1956) have demonstrated that somewhat similar indices have excellent inter-rater reliability and have considerable stability over time for a given individual. Strodtbeck (1955) and Cervin (1957) have employed similar measures for identifying the dominant member of a dyad and their results indicate considerable usefulness for interaction situation techniques.

METHOD

Subjects. The *S*s were 36 pairs of white parents who were divided into three groups of 12 couples each. The grouping of these parents was determined by the premorbid adjustment and type of illness of their biological sons. Parents in the control group had sons hospitalized for pulmonary tuberculosis with a medical history free from psychiatric illness or disturbance. These parents were selected as control *S*s because tuberculosis shares a number of important features with mental disorder. Society is antipathetic to both illnesses, and both result in prolonged hospitalization and social isolation of the patient. Moreover, it was possible to employ an identical procedure with these parents and the parents of schizophrenic patients. Parents in the Good Premorbid group (Goods) had sons hospitalized for schizophrenia with the symptom picture uncomplicated by organic factors or mental retardation. In addition, these sons had made a good adjustment prior to psychosis as indicated by a Phillips Premorbid Scale score of 15 or less (Phillips, 1953). Parents in the Poor Premorbid group (Poors) were selected by the same criteria as the Goods with the difference that their sons had made a poor adjustment prior to psychosis as indicated by a Phillips Premorbid Scale score of 16 or more. The validity of this dichotomy has been reported in detail elsewhere (Rodnick and Garmezy, 1957). The Good sons were mostly married and all had numerous friends, whereas, the Poors were all single and, typically, had been social isolates. For all groups, both parents and the hospitalized son had lived together as a family unit during the son's childhood and adolescence, and the parents were living together at the time of the study. The mean age of all sons was 28.7 with a range extending from 19.5 to 36 years. None had been hospitalized for his present illness prior to the age of 17. The groups of sons were quite comparable with respect to age and education. In addition, the Tuberculous and Good groups were nearly identical with respect to total years spent in the hospital and number of years since first hospitalization. However, the Poors tended to have been hospitalized at a younger age and had spent significantly more time in the hospital in comparison with the other two groups.

Pertinent data on the groups of *S*s (couples) are presented in Table 1. The only statistically significant differences among these groups for the characteristics appearing in Table 1 occur for number of male children. The control parents had significantly more male children than the other two groups. Although it is difficult to classify many of these parents into usual socioeconomic class categories because many were engaged primarily in farming, the total sample might best be described as lower middle class, and the groups appear comparable in this respect. With the exception of one mother, all *S*s are native Southerners.

Procedure. When a set of parents satisfied the criteria listed above, they were sent a letter from the Clinical Director or Director of Professional Services of the hospital in which their son was a patient. This was the standard method of communicating with relatives of patients, and most parents had previously been contacted in this manner by the hospital. The letter stated that *E*, a clinical psychologist, was collecting information about patients hospitalized there and that the information was considered of importance in understanding and treating the son. The recipients of the letter were asked to make an appointment with *E* at their son's hospital.

When the parents arrived, they were told that information about their son and family was desirable and might help their offspring.[2] They were then interviewed about family composition and the psychotic son's premorbid history. After this interview, the father was taken to another office and asked to complete a modified form of the Parental Attitude Research Inventory (PARI) (Schaefer & Bell, 1955). *E* then returned to the office in which the mother was waiting and explained that the subsequent part of the interview would be recorded to insure *E*'s recall of what was said. *S* was then given the following instructions:

> Mrs.—, here's what I'd like us to do now. We are interested in knowing how a mother handles her problems that come up with children when

---

[2] The data collected actually served a clinical purpose. Such data were made available to professional personnel who were working with sons of these parents.

> husband is not around. I'm going to describe some problems that come up more or less often, and I'd like you to tell me what you would do. Remember, imagine that your husband is not around to help you so that you have to handle these problems yourself.

Twelve hypothetical problem situations, adapted from a study by Jackson (1956), were then read, and S was given unlimited time to respond to each one. Three examples of these problem situations are given below:

> 1. Your twelve-year-old son gets home from school. He throws his books down, telling you what you have been teaching him is wrong and that he is not going to listen to you any more.
> 2. A gang of boys calls to your eight-year-old son to come out and play. You don't think it's good for your son to play with these boys, but now he starts to leave the house to go with them.
> 3. Your seventeen-year-old son has a chance to take a job which you are sure would be good for him. He knows you would like him to take it, but he doesn't want it. Instead, he wants to take a job which you think would not be good for him.

When the task was completed, the roles were reversed. The mother was taken to the other office to complete the PARI, and the father was given the same instructions and the same hypothetical situations. Finally, both parents were once more brought together in the office containing the tape recorder and given the following instructions:

> Both of you have told me how you would handle problems that come up with children when you are by yourselves. We are also interested in knowing how parents would handle these problems together. Imagine that these same problems came up when both of you are there. I would like both of you to agree about what you would do and tell me what that is. Be sure both of you say "agreed" when you are finished.

The tape recorder was started once more and the problem situations were again read by E. For all Ss, these hypothetical problems were read in the same order.

Throughout the interaction between the parents, great care was exercised to insure that E did not influence their behavior. Verbal reinforcement was given only once—and this came in all instances—at the end of the parent's first response during the initial interview in which each was seen individually. Furthermore, the extent and quality of E's participation in the interaction was guided by the Jackson content scoring categories (Jackson, 1955, 1956). All questions by E were designed to elicit scorable responses from the parents or to determine whether or not agreement had been reached between them.

Indices of Dominance. The dominance and conflict measures employed in the analyses were taken exclusively from the interview with both parents with the exception of the "Yielding" indices, which were derived from both the individual and joint sessions. The several indices of dominance were operationally defined as follows:

1. Speaks First—The number of times one parent spoke first in 12 situations minus the number of times the other spoke first.
2. Speaks Last—The number of times one parent spoke last in 12 situations minus the number of times the other spoke last.
3. Total—First and Last—Sum of 1 and 2 above.
4. Passive Acceptance of Solution—The number of times in 12 situations one parent passively accepted the other's solution minus the number of times the other behaved similarly. A simple unelaborated agreement in response to the spouse's solution constituted passive acceptance.
5. Total Time Spoken—One parent's total speaking time for 12 situations relative to the other's total speaking time.
6. Yielding-Maximum—The content of the responses to the problem situations was scored along the dimension of amount of pressure to conform which the parent, by his or her action, would bring to bear on the child. An ordinal scale developed by Jackson (1955) was employed for this purpose, and yielding scores were defined in terms of scale points yielded by one parent in moving toward an acceptance of the spouse's solution. For example, if the mother was very lenient, whereas the father applied great pressure to the child in the individual interviews, and together they arrived at a lenient solution, then the father was considered to have yielded to the mother. When more than one solution was offered to one problem situation, Yielding-Maximum was always based on the most severe solution offered by the parents, both in the individual and joint interviews.
7. Yielding-Minimum is identical to Yielding-Maximum except that it was always derived from the least severe solution offered.

Indices of Conflict.[3] These indices were chosen on the assumption that they would reflect the degree of respect and tolerance the Ss had for each other and would indicate the degree to which they could compromise and cooperate. They were operationally defined as follows:

1. Frequency of Simultaneous Speech—The number of occasions in the interview during which both parents spoke concurrently.
2. Duration of Simultaneous Speech—The total time in seconds during which both parents spoke concurrently.
3. Interruptions - Mother—The number of times the mother interrupted the father during the interview.
4. Interruptions - Father—The number of times the father interrupted the mother.
5. Interruptions - Total—Sum of 3 and 4 above.
6. Disagreements and Aggressions - Mother—The number of times the mother disagreed with or aggressed against the father during the interview.
7. Disagreements and Aggressions - Father—The number of times the father disagreed with or aggressed against the mother during the interview.
8. Disagreements and Aggressions - Total—The sum of 6 and 7 above.

---

[3] For further information see Farina (1958).

9. Failure to Agree—The number of situations for which the parents failed to arrive at a mutually acceptable solution.

10. Verbal Activity—The total amount of time during which the mother and father talked.

While most of the indices comprising the conflict variable appear to have considerable face validity, it may not be apparent why the index Verbal Activity was also used as a measure of conflict. Since the instructional set given to the parents in the interaction interview stressed that they arrive at a joint solution, *S*s continued to speak as long as they remained in disagreement. To test the assumption that Verbal Activity could serve to reflect the difficulty experienced by the parents in reaching agreement or the extent of the disagreement between them, total time talked by both parents in situations where agreement was reached was compared to total time talked in those situations in which they failed to achieve a common solution. Since the difference was in the expected direction and highly significant ($t = 4.76$, $df = 22$, $p < .001$) it may be concluded that *S*s do tend to talk longer in situations where they do not agree.

Reliability of Indices. Four cases were chosen for independent rating by an advanced graduate student in psychology[4] and two of these four cases were rerated by *E* without reference to his previous ratings. The resulting inter- and intrarater correlation coefficients were very similar. For four indices, Speaks First, Speaks Last, Failure to Agree, and Passive Acceptance of Solution, there was virtually perfect agreement. For these, one disagreement occurred in 360 possibilities. For the balance of the indices, the average Spearman *rho* value for inter- and intrarater reliabilities combined was +.89 with a range from +.74 to +.96. Although there are variations from index to index, each of these indices appeared sufficiently reliable to warrant its retention.

RESULTS

The main purpose of this study was to test the hypothesis that there is an association between the sex of the dominant parent and the premorbid adjustment of the schizophrenic son. The analysis of the relevant data was done by deriving indices from the mother and father scores in such a way that negative scores indicated mother dominance and positive scores father dominance. For example, for Speaks First, the number of times the mother spoke first in 12 problem situations was subtracted from the number of times the father spoke first. Table 2 presents the means and standard deviations for each group on each index. Increasingly high positive numbers indicate increasing degrees of father dominance, whereas mother dominance is similarly represented by negative numbers. In the case of Total Time Spoken the mother's score was divided into the father's score so that an index greater than unity indicates father dominance and one less than unity indicates mother dominance. In brief, Table 2 shows that mother dominance was most marked for the Poor group, that the controls were intermediate, and that the Good group was characterized by father dominance.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF GROUPS ON DOMINANCE INDICES[a]

	Controls	Goods	Poors
Speaks First			
Mean	−.5	+5.7	−2.2
SD	9.1	6.3	7.0
Speaks Last			
Mean	−1.0	+3.4	−2.7
SD	5.0	4.0	5.7
Total First and Last			
Mean	−1.5	+9.2	−4.8
SD	10.6	7.1	9.7
Passive Acceptance of Solution			
Mean	−.5	+3.1	−1.2
SD	4.4	3.2	1.9
Total Time Spoken[b]			
Mean	.72	2.0	.72
SD	1.9	3.2	1.3
Yielding-Maximum			
Mean	+.12	+1.3	−.8
SD	2.4	1.1	1.7
Yielding-Minimum			
Mean	−.13	+.8	−1.0
SD	1.9	1.6	1.6

[a] + indicates father dominance; − indicates mother dominance.

[b] A score greater than 1 indicates father dominance; a score below 1 indicates mother dominance.

The significance of the difference among these three sets of means was evaluated by a nonparametric test suggested by Jonckheere (1954). The *p* values yielded by this test are shown in the first column of Table 3. All differences are significant at at least the .06 level of probability. A further analysis to determine the probability that any two groups were selected from the same population was made by means of the Mann-Whitney *U* test. The *p* values yielded by this series of tests are presented in the last three columns of Table 3. On this table, *p* values greater than .05 are listed as nonsignificant (*NS*). For all indices, the Good and Poor groups differed from each other at better than the .01 level of significance in the expected direction. All of the indices

---

[4] Appreciation is expressed to Richard Dunham for this work.

TABLE 3

LEVELS OF SIGNIFICANCE OF DIFFERENCE AMONG AND BETWEEN GROUPS FOR DOMINANCE INDICES

Index	Among Groups	Between Groups		
		Controls vs. Goods	Controls vs. Poors	Goods vs. Poors
Speaks First	.06	*NS*	*NS*	.01
Speaks Last	.019	.025	*NS*	.01
Total First & Last	.004	.025	*NS*	.001
Passive Acceptance of Solution	.005	.025	*NS*	.001
Total Time Spoken	.0002	.01	*NS*	.001
Yielding-Maximum	.005	*NS*	*NS*	.001
Yielding-Minimum	.004	*NS*	*NS*	.01

in the Good-Control group comparison reveal that the Good group was characterized by more father dominance, and four of the seven indices employed reached a statistically significant difference. Six of the seven dominance indices indicate that the Poor group was more mother-dominated than the Control group, while the seventh index was identical in the two groups. In this last comparison, none of the indices reach the .05 level of significance, however.

The first three columns of Table 4 present the means and standard deviations obtained by the three groups on the ten conflict indices. For nine of these ten indices, extent of con-

TABLE 4

MEANS, STANDARD DEVIATIONS, AND LEVELS OF SIGNIFICANCE FOR GROUPS ON CONFLICT INDICES

	Means and *SD*s			Levels of Significance			
	Control	Good	Poor	Control vs. Total Schizophrenia	Control vs. Good	Control vs. Poor	Good vs. Poor
Frequency of Simultaneous Speech				*NS*	*NS*	*NS*	*NS*
Mean	5.4	7.5	11.4				
SD	4.3	7.5	10.2				
Duration of Simultaneous Speech				*NS*	*NS*	*NS*	*NS*
Mean	4.8	9.0	13.8				
SD	3.6	12.2	16.5				
Interruptions—Mother				**	*NS*	**	*NS*
Mean	1.3	3.0	5.6				
SD	1.5	3.2	4.1				
Interruptions—Father				*	*NS*	*	*NS*
Mean	1.2	2.8	4.7				
SD	1.9	3.3	5.3				
Interruptions—Total				**	*NS*	**	*
Mean	2.6	5.8	10.3				
SD	2.4	6.3	9.1				
Disagreements and Aggressions—Mother				*NS*	*NS*	*NS*	*
Mean	1.7	1.5	4.2				
SD	1.5	2.3	5.1				
Disagreements and Aggressions—Father				*NS*	*NS*	*	*NS*
Mean	.8	1.2	2.4				
SD	1.0	1.6	2.6				
Disagreements and Aggressions—Total				*NS*	*NS*	*	*
Mean	2.5	2.8	6.7				
SD	2.3	3.3	8.0				
Failure to Agree				*NS*	*NS*	*	*NS*
Mean	.9	1.0	1.9				
SD	1.3	1.2	1.4				
Verbal Activity				*	*NS*	*	*NS*
Mean	337	424	556				
SD	106	236	295				

* $p < .05$.
** $p < .01$.

flict was least in Controls, greatest in Poors, and intermediate in the Goods. Only for the 10th index, Disagreements and Aggressions—Mother, did the Control group obtain a somewhat higher score (indicative of greater conflict) than the Goods, although for this index, too, the Poors' score was the highest. In the analysis of these data, each group was compared with every other group, and the Total Schizophrenic group, the Goods and Poors combined, was compared to the Control group. The Mann-Whitney *U* test was employed for these analyses. The resulting *p* levels are indicated in the last four columns of Table 4. The tests employed are one-tailed with the exception of the Good-Poor comparison, where the tests are two-tailed because no prediction of differences in conflict behavior had been made.

The comparisons most relevant to the hypothesis that parents of schizophrenic patients experience more conflict in their interactions than controls are those between the Control and Total Schizophrenic groups. Since all 10 indices were in the expected direction, and 4 of these reached a *p* level of less than .05, the hypothesis appears to be supported. The other comparisons reveal that the Poor group parents experienced more conflict in their interaction than either the Control or Good groups. All 10 indices were greater for the Poors than for the Controls, and 7 reached a statistically significant level. Similarly, all 10 indices were higher for the Poor than for the Good groups, and three reached statistical significance. The Good-Control groups comparisons failed to reveal any statistically significant differences for any index, although 9 of the 10 were higher for the Goods than for the Controls.

A test was made of the hypothesis that the interactions of parents of schizophrenic patients are characterized by dominance of one relative to the other. The dominance indices for these analyses were derived from the mother and father scores so that increasing magnitude of the score indicated an increasingly unequal dominance relationship between the parents. For example, for Speaks First, the submissive member's score, regardless of sex, was always subtracted from the dominant member's score. The index scores were very similar for the two groups and there was no observable directional trend. Four of the seven indices were higher for the Control group, and three were higher for the Total Schizophrenic group.

The PARI scores were very similar for all groups. Neither for mothers nor for fathers could any directional trends be observed on any of the nine scales administered.

DISCUSSION

The results of this investigation revealed a marked relationship between sex of the dominant parent and level of premorbid adjustment achieved by the schizophrenic son. The interactions of parents of good premorbid patients were characterized by father dominance whereas the mother was dominant in the case of parents of poor premorbid patients. However, this correlation does not permit one to assume that such parental behaviors were the cause of the offspring's behavior since this study shares the problems inherent in any retrospective study.[5] Clearly, one can posit alternative explanations for the relations between parental and patient behaviors. For example, it is conceivable that the son's role in the family in the past may have given rise to current parental interactions patterns or that parental awareness of the son's status and their own relationships to the hospital may have served to modify their situational test behavior. Nevertheless, the traits displayed in a situational test such as the one employed in this study may have considerable stability over time and therefore may be representative of earlier behaviors in the family setting. For example, Naka (1948) has demonstrated that certain personality traits are persistent in most individuals over a 15-year time span. Her *S*s were children of 2 and 17 years of age, however.

If it is assumed that the parental behaviors as measured in this study were characteristic of the parents from the time the patient was an infant, how might these have been significant in shaping the offspring's personality? The good premorbid schizophrenic patients displayed more customary adult male behaviors prior to psychosis than was true of the poor premorbid patients. The family environment in which the good premorbid schizophrenic develops may make it easier for him to achieve a somewhat higher level of maturity and to manifest more adequate male behavioral patterns relative to the poor premorbid schizophrenic patient. Although the good premorbid may still be exposed to such pathogenic family influences as are generated by a tyrannical father and a weak, vacillating mother, the father does appear to assume dominance in the family. Consequently, she is in a position to reward the son for imitating his behavior and to serve as a model, even if a faulty one, for male identification. The poor premorbid schizophrenic, on the other hand, appears to have grown up in a family where the mother exercises control and where the father does not serve as a reasonably adequate male figure with whom the patient can identify. Even in instances where the son makes tentative efforts to identify with the father, the antagonism between the parents may inhibit an effective identification as Mowrer (1950, p. 596) has pointed out. These findings appear to have implications for theories of identification and sex typing, since the behavioral differences between the good and poor premorbid schizophrenic patients are relevant to these concepts.

---

[5] These and related problems of retrospective studies will be discussed at length in a forthcoming publication (Garmezy, Farina, & Rodnick, 1960).

The findings concerning conflict appear to support the reports by many writers that relationships between parents of schizophrenic patients are characterized by hostility and contentiousness. If their behavior in the situational test can be used to infer their more general behavior, these parents do not appear to be able to cooperate with one another. Impatience with each other and a mutual lack of respect and regard are indicated by frequent interruptions and unwillingness to compromise in the task. This discord may involve the schizophrenic offspring according to Lidz, Cornelison, Fleck, and Terry (1957b), who report that he is often the focus and battleground of parental conflict. Since the stimulus situations employed in this study dealt with problems of disobedience and misbehaviors of children, they may have served as effective cues for eliciting the conflictful parental interactions.

Excessive dominance by one parent, particularly dominance by the wife over the husband, has often been described as the modal parental pattern in families of schizophrenic patients. The results of the present investigation, however, suggest that the premorbid adjustment of the schizophrenic son should be considered in describing the parental dominance relationship. In fact, when the sex of the parent and the premorbid adjustment of the son are disregarded and the Goods and Poors combined are compared to the Control group, no differences in extent of dominance of one parent over the other can be discerned in these samples. Furthermore, since the role the parents play in the child's acquisition of sex-appropriate behaviors appears to vary with the sex of the child (Sears, Pintler, & Sears, 1946), and since the process of identification may be different for females, the sex of the schizophrenic offspring should also be considered in studying the role of possible family influences in the development of a schizophrenic disorder.

SUMMARY

The major purpose of the present study was to investigate a reported association between sex of the dominant parent and the adequacy of premorbid adjustment of the schizophrenic son. The investigation was carried out by means of a situational test so designed and structured that indices of dominance could be obtained. Such indices were length of time spoken, relinquishing one's position in favor of the spouse, etc. Paternal dominance was found to be associated with good premorbid adjustment of the schizophrenic son whereas in the case of those patients whose premorbid adjustment was poor, the mothers were dominant. Because of these differences in parental behaviors, it may have been somewhat easier for the good as compared to the poor premorbid son to learn the more usual adult male behaviors and to achieve a somewhat higher level of maturity.

Extent of conflict present in the parental interactions was inferred from such measures as number of interruptions, number of disagreements, etc. The parents of the schizophrenic patients displayed more conflict than the Control parents, and this conflict was particularly marked in the parents of the poor premorbid schizophrenic sons. The results of the present investigation suggest that the premorbid adjustment and sex of the schizophrenic offspring need to be considered when investigations of the familial role in schizophrenia are undertaken.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955.

CERVIN, V. Relationship of ascendant-submissive behavior in dyadic groups of human subjects to their emotional responsiveness. *J. abnorm. soc. psychol.*, 1957, **54**, 241–249.

ELLISON, E. A., & HAMILTON, D. The hospital treatment of dementia praecox. Part II. *Amer. J. Psychiat.*, 1949, **106**, 454–461.

FARINA, A. Patterns of role dominance and conflict in the interaction of parents of schizophrenic patients. Unpublished doctoral dissertation, Duke Univer., 1958.

FRAZEE, H. E. Children who later became schizophrenic. *Smith Coll. Stud. soc. Wk.*, 1953, **23**, 125–149.

FRIEDLANDER, D. Personality development of 27 children who later became psychotic. *J. abnorm. soc. Psychol.*, 1945, **40**, 330–335.

FROMM-REICHMAN, FRIEDA. Notes on the development of treatment of schizophrenia by psychoanalytic psychotherapy. *Psychiatry*, 1948, **11**, 263–273.

GARMEZY, N., FARINA, A., & RODNICK, E. H. The structured situational test: A method of studying family interaction in schizophrenia. *Amer. J. Orthopsychiat.*, in press.

GERARD, D. L., & SIEGEL, J. The family background in schizophrenia. *Psychiat. Quart.*, 1950, **24**, 45–73.

JACKSON, P. W. Verbal solutions to parent-child problems and reports of experience with punishment. Unpublished doctoral dissertation, Columbia Univer., 1955.

JACKSON, P. W. Verbal solutions to parent-child problems. *Child Developm.*, 1956, **27**, 339–349.

JONCKHEERE, A. R. A distribution-free K-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

KOHN, M. L., & CLAUSEN, J. A. Parental authority behavior and schizophrenia. *Amer. J. Orthopsychiat.*, 1956, **26**, 297–313.

LIDZ, T., CORNELISON, A. R., FLECK, S., & TERRY, DOROTHY. The interfamilial environment of the schizophrenic patient. *Psychiat.*, 1957, **20**, 329–342. (a)

LIDZ, T., CORNELISON, A. R., FLECK, S., & TERRY, DOROTHY. The interfamilial environment of schizophrenic patients: II. Marital schism and marital skew. *Amer. J. Psychiat.*, 1957, **114**, 241–248. (b)

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

MATARAZZO, J. D., SASLOW, G., & GUZE, S. B. Stability of interaction patterns during interviews: A replication. *J. consult. Psychol.*, 1956, **20**, 267–274.

MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald , 1950.

NAKA, PATRICIA N. Shirley's babies after fifteen years: A personality study. *J. genet. Psychol.*, 1948, **73**, 175–186.

OPLER, M. K. Schizophrenia and culture. *Scient. Amer.*, 1957, **197**, 103–110.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

REICHARD, SUZANNE, & TILLMAN, C. Patterns of parent-child relationships in schizophrenia. *Psychiatry*, 1950, **13**, 247–257.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109–183.

ROSEN, J. N. *Direct analysis*, New York: Grune & Stratton, 1953.

SCHAEFER, E. S., & BELL, R. Q. Parental attitude research instrument. (PARI). Unpublished manuscript. Lab. Psychol., Natl. Inst. Ment. Hlth., 1955.

SEARS, R. R., PINTLER, M. H., & SEARS, P. S., Effect of father separation on preschool children's doll play aggression. *Child Developm.*, 1946, **17**, 219–243.

STRODTBECK, F. L. Husband-wife interaction over revealed differences. In A. P. Hare, E. F. Borgatta, & R. F. Bales (Eds.), *Small Groups*. New York: Knopf, 1955, 464–472.

TIETZE, TRUDE. A study of mothers of schizophrenic patients. *Psychiatry*, 1949, **12**, 55–65.

(Received December 8, 1958)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 39-46

THE VALIDITY OF HYPNOTICALLY INDUCED VISUAL HALLUCINATIONS[1]

HAROLD WAYNE UNDERWOOD[2]
University of Illinois

STUDIES on the influence of hypnotic suggestion on memory, physical strength, emotional and physiological responses, etc., have established that many hypnotic phenomena involve more than simply acting. One kind of problem that has been particularly formidable, however, has been that of determining to what extent hypnosis can influence perception. Here, the problems involved in studying perception, such as difficulty in insuring that the overt response is an accurate reflection of a percept, are added to problems in studying hypnosis.

The purpose of this study was to test the validity of hypnotic positive[3] and negative[4] visual hallucinations. Operationally, the problem is to determine whether or not there is a percept resulting from a hypnotically suggested visual stimulus, which is similar to the percept resulting from the objective stimulus, and whether "suggesting away" a stimulus produces changes in a percept similar to those produced by objective removal of the stimulus. The primary difficulty in such an investigation is that of insuring that the hypnotic subject (*S*) is not merely faking or pretending to see a suggested stimulus.

METHOD

To test the validity of hypnotic hallucinations, three optical illusions were used in which a field distorts a superimposed figure. For example, in one illusion, a series of inverted Vs (field) superimposed on a square (figure), make the sides of the square appear to widen out toward the top (see Fig. 5). If hypnotic hallucinations are valid, then the positive hallucination of a field on a real figure should cause appropriate figure distortion and likewise, the negative hallucination of an actual field should remove the distortion of the figure. The distortion of figures was measured in terms of quantity as well as simply presence or absence, thereby providing a measure of the degree of effectiveness of the hypnotic suggestion in contrast to objective stimulation. Unfamiliar illusions were used, which were designed so that distortions could not be inferred without perceiving them, unless the *S* had a fairly precise knowledge of the illusions.

Subjects

Hypnotic *S*s were selected by screening approximately 195 *S*s (mostly graduate students) for hypnotic susceptibility by attempting to hypnotize them. Group hypnosis was used on about 100 of those *S*s, in five different groups, and the remaining 95 were seen in individual sessions. *S*s who showed susceptibility in either a group or individual session were seen privately for several sessions to determine whether or not they would be capable of achieving sufficient trance depth to hallucinate. Twelve *S*s were chosen as being suitable for the study on the basis of seeming to have the most vivid hallucinations, while lacking in any gross visual defects or any background that might have made them familiar with the illusions. They all responded positively to trance depth tests of eyelid and arm catalepsies, arm levitation, hand anesthesia, selective amnesia, positive and negative tactual, auditory and visual hallucinations, and posthypnotic suggestions, including trance induction by signal. Also, they were all somnambulists; that is, they could walk, talk, and have their eyes open while remaining in the trance.

Even within this highly select group, six *S*s showed imperfections in their trance performance. Three reported hallucinations that were sometimes "ghost-like" or fluctuating in vividness, and they also reported that hallucinating required considerable effort. Three other *S*s reported very vivid hallucinations, but were obviously not hallucinating with ease. There was much squinting, tear secretion, fatigue and other signs of excessive effort. The remaining six *S*s reported constantly vivid hallucinations (indistinguishable from reality) which were apparently produced with ease.

Since past research has suggested that extremely deep trances are necessary for this type of research (Erickson 1938a) those six *S*s who showed trance imperfections were delegated to a separate group called the "deep group." The six who showed flawless performance were called the "very deep group."[5]

---

[1] This paper is based on a PhD dissertation at the University of Illinois. The author wishes to express gratitude to William M. Gilbert for his generous advice and guidance as chairman of the doctoral committee.

[2] Now at the VAH, Danville, Illinois.

[3] Positive hallucination is the perception of a suggested stimulus, which is in fact not present.

[4] Negative hallucination is the failure to perceive an actual stimulus which is suggested to be absent.

The control group of six *S*s was picked to match the hypnotic groups roughly in age and educational background. No member was to have ever been hypnotized. It was composed of two female and four male *S*s with a median age of 26.5: three graduate students, one undergraduate, one with a master's degree, and one with a high school education.

The deeply hypnotized group had one male and five female *S*s with a median age of 27.5. Two *S*s had bachelor's degrees, one *S* was an undergraduate, one a business college student, and two had a high school education. The very deeply hypnotized group consisted of three male and three female *S*s with a median age of 22.5: three graduate students, one undergraduate, a nurse, and one with business college training. None of the graduate students was in the field of psychology, and no *S* had any training regarding illusions beyond that obtained in an introductory psychology course.

Apparatus

The illusions were presented to the *S*s by projecting the figures from behind onto 8" x 10" translucent glass screens on which the fields were drawn. A 35 mm. slide projector and frame holders for the screens were mounted on a table with a shield surrounding the screen to block the experimenter (*E*) from the view of *S*.

Illusion A (Osgood, 1953), illustrated in Figs. 1–3, uses a "spoked wheel" field with a square as a figure. The distortion of the square depends on its size and position within the field. When off center of the field to the left (Fig. 1), it is commonly seen as bulging on the right, when directly over center (as in Fig. 2), it has no distortion, and when small and totally to the left of center (Fig. 3), it commonly appears concave on the right and narrowing to the left.

Illusion B (Osgood, 1953), shown in Fig. 4, employs a field of concentric circles which commonly causes a "square" (with slightly convex sides) to appear to have slightly concave sides.

Illusion C (Berliner, A. & Berliner, S., 1948; Henriksson, 1955; Hofmann & Bielschawsky, 1909), which has a field of inverted Vs, is illustrated in Fig. 5 with a square as a figure. In addition to the square 11 trapezoids were used as figures for this illusion (shown in Fig. 6). The effect of the field on each figure is to make the sides seem to angle out to a wider top than is objectively accurate. Even though the field had the general effect of widening each figure toward the top, there were differences among the figures in amount of widening because of the slightly different angle at which the field lines crossed each of the sides of the figures. Since *S*s could hardly infer this differential illusory effect among the figures, it seemed promising as a control for faking.

Illusion A was used to test the positive hallucination (+H), Illusion B to test the negative hallucination (−H), and Illusion C to test both the +H and −H. A special problem existed in testing the +H in that it was necessary to insure that the *S* was hallucinating the field in the proper position over the figure. To accomplish this, minimal parts of the field were presented on the screen, and the *S* merely had to hallucinate the missing portions. These minimal parts of the field have been called "reference patterns." The reference patterns for Illusions A and C are illustrated in Figs. 7 and 8 respectively, with example figures placed properly.

*S*s reported perceptions by a matching technique. A set of figures on a card was placed below the screen on which illusions were presented, and *S* was required to select from that set a figure matching the one observed on the screen. One set of 18 figures was used for Illusions A and B (Fig. 9). From this set there was only one correct matching figure for each illusory figure, and hence responses could be scored as correct or incorrect. A second set for Illusion C (illustrated in Fig. 6) consisted of duplicates of the illusion figures. It may be seen that they formed a progression of figures. The amount of distortion could be reported rather than simply the type.

Conditions

All *S*s reported perceptions of figures under the five following conditions:

1. Positive hallucination (+H)—figures were shown on the reference pattern with the suggestion that the pattern was filled in to be exactly like the field which he could see on a second screen beside the reference pattern.
2. Negative hallucinations (−H)—figures were shown superimposed on the field with the suggestion that the field was absent.
3. Reference pattern—figures were shown on the reference pattern with no hallucination suggestion. This control was necessary to insure that any distortions of figures in the positive hallucination condition were not caused by the reference pattern alone rather than the hallucinated part of the field.
4. Objective—figures were shown on a blank screen.
5. Illusion—figures were shown on the field. The objective and illusion conditions formed a basis for evaluating the hallucination conditions. If hallucinations are valid, the responses to the +H condition should differ from the objective while approaching or being like the illusion condition, and the opposite should be true for the −H.

In determining the order in which illusions and conditions should be presented, the primary concern was to avoid giving *S* any information at one time that could be used in faking responses later. Obviously, for example, it was necessary to have the +H condition precede the illusion condition.

The order of presentation of the conditions and Illusions A, B and C was: A–reference pattern, A–(+H), B–illusion, B–(−H), C–objective, C–reference pattern, C–(+H), C–(−H), C–illusion, A–illusion, and B–objective with the exception that C–objective and C–reference pattern were reversed with alternate *S*s in all groups.

---

[5] The very deep group was selected first and run through the experiment. When the quest for further *S*s failed to produce *S*s with flawless trance performance, the deep group was formed and run through the experiment.

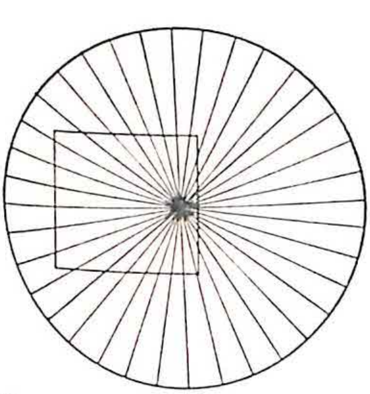

FIG. 1. Illusion A, Illusion Fig. 1.

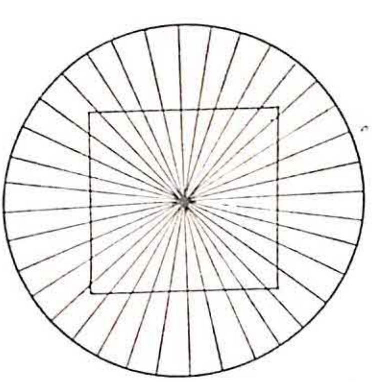

FIG. 2. Illusion A, Illusion Fig. 2.

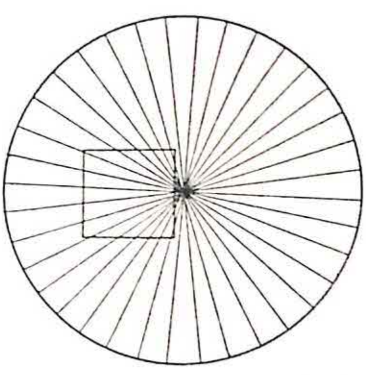

FIG. 3. Illusion A, Illusion Fig. 3.

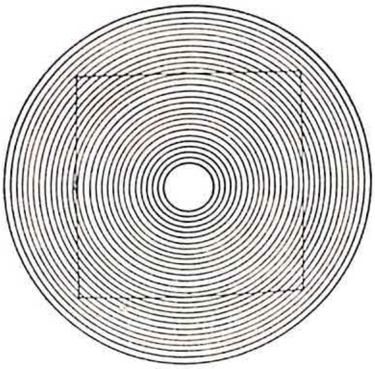

FIG. 4. Illusion B.

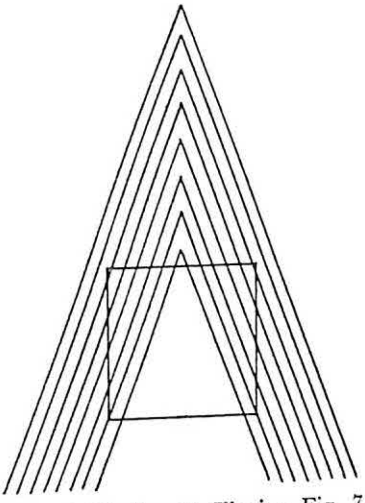

FIG. 5. Illusion C, Illusion Fig. 7.

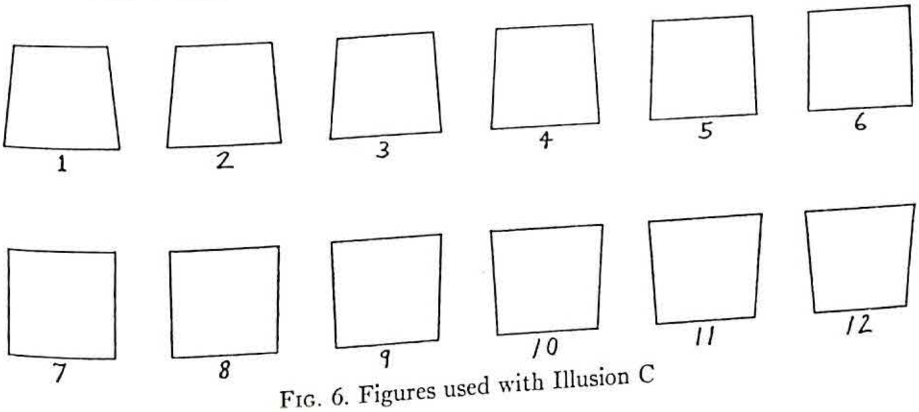

FIG. 6. Figures used with Illusion C

The three figures of Illusion A were always presented in order as numbered. With Illusion C, all the figures (1 through 12) were presented under objective and reference pattern conditions, Fig. 3 through 8 were presented for the hallucination conditions, and Fig. 1 through 8 for the illusion condition. They were presented in different random order for each condition and the same order for all *Ss*.[6]

Hypotheses

On the basis of past research (Benussi, 1927; Erickson & Erickson, 1938b) and a pilot study, it was hypothesized that in very deeply hypnotized *Ss*, the positive hallucination (+H) of the field of an illusion would result in reports of appropriate distortion in the superimposed figure and the negative hallucination

---

[6] Figures 9 through 12 were left out when distortion was anticipated since there would not be a matching figure to report. Figures 1 and 2 were left out of the hallucination conditions since the direction of the illusory distortion might be inferred from leaving figures off from only the upper end of the scale.

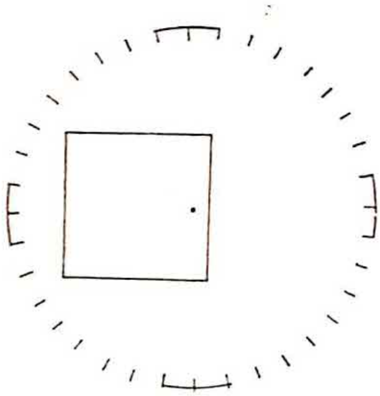

FIG. 7. Reference pattern for Illusion A with Illusion Fig. 2.

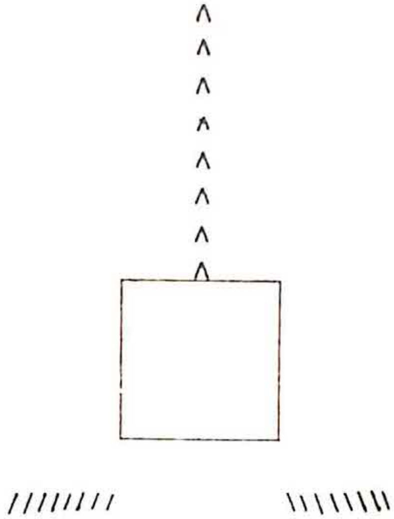

FIG. 8. Reference pattern for Illusion C with Illusion Fig. 7.

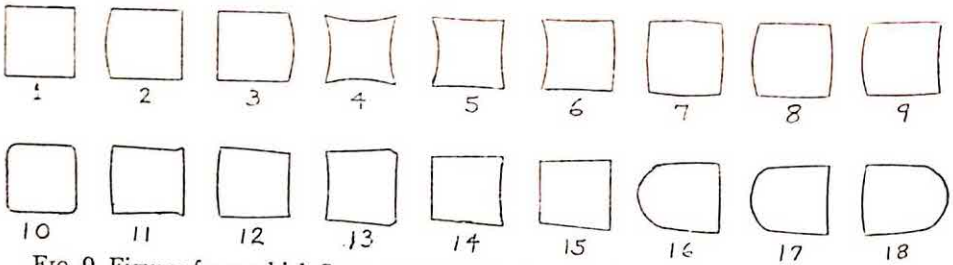

FIG. 9. Figures from which S was to select matches for figures of Illusions A and B.

(−H) of the field would result in reports of a reduction or removal of the figure distortion. Control *S*s were expected to show less or no such hallucinatory effects. After data were collected from those two groups, it was hypothesized that the group of deeply hypnotized *S*s would show either no such hallucinatory effect or less effect than that shown by the very deeply hypnotized group.

Procedure

Hypnotic *S*s were trained to reach sufficient trance depth for their group (see *Subjects*, above). The eye fixation and hand levitation techniques were used, the former more commonly. Three to five one-hour sessions were typically required to achieve sufficient depth in suitable *S*s. Control *S*s were given no training prior to the experimental session.

At the experimental session *S* was seated in a chair about three feet from the screen. Hypnotic *S*s were hypnotized by the technique suitable to them and trance deepening suggestions were given for 5 to 10 min. depending on the particular *S*. They were tested for reports of −H and +H with stimuli unrelated to the illusions and then went through all the conditions while hypnotized. The control *S*s simply went through all conditions in the nonhypnotized state.

Instructions for all *S*s on objective, reference pattern and illusion conditions were exactly the same. *S*s were simply asked to match the projected figures with those below the screen. It was specified with Illusions A and B in all conditions that they were to pick figures showing the kind of shape, though size and amount of curve or angle might be different. With the hypnotic *S*s in +H condition the experimenter (*E*) instructed that the reference pattern was going to gradually fill in to become exactly like the field (an actual field was beside the reference pattern) as *E* counted to 10 and it would be completely filled in by the time 10 was reached. The *S* was also told that his right hand and forearm would lift as lines filled in, until when the designs were exactly the same, his arm would be pointing straight up. Having carried out those steps, *E* then instructed that the lines were to remain completely filled in just like the field until otherwise instructed; however, if for any reason the lines should fade or change in any way, his hand was to lower. He was also to make judgments about the figures on the screen only when the lines were completely filled in. The figures were then shown on the hallucinated field, and then the hallucination was removed. Similar instructions with appropriate modification were used for −H with the exception that *S*s were told not only that the screen would become blank, but that they would forget what had been on it before.

Instructions to control *S*s on the −H and +H conditions were the same as for hypnotic *S*s, except that they were told to *imagine* the lines being filled in or absent and they were to *guess* the forms the projected figures would appear to have, if the suggested stimulus condition actually existed. Throughout the instructions, there were minor differences from the wording for the hypnotic groups, which were consistent with *S*'s imagining rather than *E* asserting actual stimulus changes.

After going through all conditions, the control *S*s were questioned about prior knowledge of the illusions and the effectiveness of imagining the stimuli. Hypnotic *S*s were questioned in the trance about prior knowledge of illusions, clarity of hallucinations, and whether they ever reported percepts different from what they actually saw. They were instructed to awaken and the same questions were asked again.

RESULTS

Illusion A

In the +H condition (Illusion A was not used to test the negative hallucination), the only *S*s reporting correct distortions were in the very deep trance group. In that group two of the six *S*s reported correct distortions on illusion Fig. 1 and also reported perceiving concaveness, though missing the narrowing on Fig. 3. A third *S* in that group also reported the same perceptions, but then changed his mind stating that all three figures looked like squares. Regarding Fig. 2 in which there is no noticeable illusory distortion, the number of *S*s reporting the correct percept (square) in the control, deep trance, and very deep trance groups were respectively three, six, and four.

All *S*s reported seeing squares (i.e. no distortion) in the reference pattern condition and the illusory effect in the illusion condition.

Statistically, neither the very deep trance nor the deep trance group is significantly different from the control group on any of the conditions (Fisher's exact probability test). Thus, the results suggest a tendency toward a +H effect only in the very deep trance group, which is statistically insignificant.

Illusion B

No *S* gave the correct response to Illusion B under the −H condition. Since the +H condition was not used with this illusion, evidence here pertains only to −H, and is negative. Reported perceptions on the objective condition were veridical and the illusory effect was reported on the illusion condition.

Illusion C

Results from Illusion C are presented in Fig. 10 and Table 1.

Regarding Fig. 10, it should be remembered that plus error is normal illusory error. First with respect to the −H condition, it may be seen that there is essentially no evidence for negative hallucinations (−H responses do not deviate from illusion toward objective responses), a finding consistent with results from Illusion B. There is, however, clearly a trend in the very deep trance group indicating a tendency toward the effects expected with positive hallucinations (+H responses deviate from objective toward illusion responses), and no such effect in the control or deep trance group. Considering +H effect in terms of number of *S*s showing it, one *S* in the control group, no *S* in the deep group, and four *S*s in the very deep group reported figure distortion in the correct direction. Figure 10 also shows that the control group tended to guess that the illusion caused distortion opposite from that actually produced, as indicated by the error scores under both the +H and −H conditions.

In order to evaluate these and other trends, a three dimensional analysis of variance was performed. As indicated in Table 1 with all 18 *S*s, the only significant *F* value results from variance introduced by conditions. The index to the hallucinatory effect is the group by condition interaction, and the obtained *F* ratio would be expected by chance slightly more than 10% of the time. The responses of one *S* from the very deep group are sufficiently deviant from those of all other *S*s that a con-

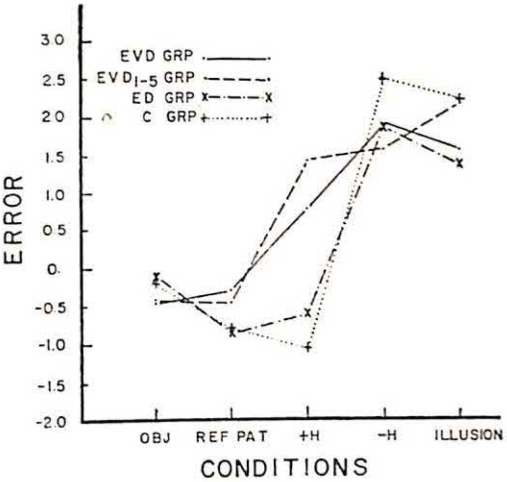

FIG. 10. Mean error per figure judgment with Illusion C, Fig. 3–8, in the very deeply hypnotized group (Evd), very deeply hypnotized group excluding deviant *S* (Evd 1–5), deeply hypnotized group (Ed) and control group (C). Error is in figure units.

TABLE 1
ANALYSIS OF VARIANCE OF JUDGMENTS OF ILLUSION C FIGURES

Source	Mean Square	*df*	*F*
Subjects	4.482	17	
Groups	6.236	2	1.468
Subjects within Groups	4.247	15	
Conditions	177.467	4	28.522*
Figures	1.193	5	
Conditions by Figures Interaction	1.999	20	1.126
Groups by Conditions Interaction	10.638	8	1.710
Groups by Figures Interaction	2.415	10	1.730
Groups by Figures by Conditions Interaction	1.257	40	1.116
Pooled *S*s by Conditions within Groups Interaction	6.222	60	
Pooled *S*s by Figures within Groups Interaction	1.395	75	
Pooled *S*s by Conditions by Figures within Groups Interaction	1.126	300	

* Significant at .001 level.

siderably more complete description of the results is gained by presenting analyses of both the complete data and data with that case excluded. This *S* was unique in reporting the opposite from correct perceptions on +H, −H and illusion conditions on all three illusions. His responses to the reference pattern condition on Illusion C were very erratic, some being markedly in error in both directions and others fairly accurate. Considering all 12 figures of Illusion C, he increased the variance of the very deep group on reference condition from 1.5 to 33.9. Poststudy questioning revealed that a few years prior to the study he had been hospitalized for a hysterical reaction. By excluding the deviant *S*, the *F* ratio from group by condition interaction increases to the point that it would be expected by chance less than 5 times in 1000. Also the variance due to groups gains considerable significance, barely missing the .05 level. Thus, there is no conclusive evidence for a group by condition interaction, but a tendency toward such an effect is present, and the tendency is given further weight when the inordinate contribution of one *S* is taken into account.

To further isolate the source of variance in the group by condition interaction, *t* tests were performed, contrasting the groups on the various conditions. Among the group comparisons, the only significant *t* ratios are found on the +H condition. The contrast between the control group and the entire very deep group reaches the .05 level of significance; when the deviant *S* is excluded, the contrast reaches the .005 level (one alternative *t* tests).

When the differences between conditions are tested over all three groups together, each condition is significantly different from each other condition, except for the contrasts of the objective vs. positive hallucination, reference pattern vs. positive hallucination, and negative hallucination vs. illusion. The significance levels by the sign test are greater than .001 except for the objective vs. reference pattern contrast which is .004.

When the differences between conditions are tested within groups, there are three reversals to the results over the three groups together. The objective vs. reference pattern contrast is insignificant by the *t* test in the very deep group, though it remains significant if the deviant *S* is excluded. And objective vs. positive hallucination and reference pattern vs. positive hallucination contrasts are significant in the very deep group excluding the deviant *S* (.05 and .025 respectively by one alternative *t* tests).

One peculiar outcome was that figures were judged to be significantly more narrow toward the top in the reference pattern condition than in the objective condition.

Poststudy Questioning

No subject reported having seen any of the illusions prior to the study and all answers were the same in and out of the trance with the exception that only while hypnotized, one *S* from the very deep group reported possibly having seen Illusion A (none of his +H responses were correct, however, on that illusion). All in the control group reported being unable to imagine lines appearing or disappearing. In the very deep group, all *S*s reported complete disappearance of lines during the −H. Four of those *S*s reported complete +Hs and the other two reported a slight fading of lines at times during the +H, in spite of having failed to report such fading at the time. These were the two *S*s who only showed hallucinatory effects part of the time. In the deep group, three *S*s reported complete −Hs and +Hs.

and the remaining three *S*s reported blurring of lines beyond recognition during the −H and faded or incomplete +Hs.

There was thus no evidence that *S*s had any usable prior information about the illusions. The reports of positive hallucinations were generally reflected in the objective measures, although there were several cases in which claims of hallucinations were not borne out. Even though *S*s generally reported having negative hallucinations, objective indication of their occurrence was lacking.

DISCUSSION AND CONCLUSIONS

The evidence is quite conclusive that reported negative hallucinations failed to alter perception of the field, to a sufficient extent to lessen measurably its distorting effect on the figure. There is, in fact, no evidence that there was any perceptual alteration at all accompanying the −H. One may not conclude that there was none, however, since perceptual alteration without influence on the illusion is conceivable. Relative to this question, evidence exists which indicates that illusory processes occur both retinally and cortically (Motokawa, 1950). If the hypnotic −H is cortical in locus, then the field processes might have already distorted the figure in the retina at a point prior to the locus of hypnotic influence.

There is conclusive evidence that the field +H did not have the same effect on the *S*s as did objective field stimulation. Even when *S*s were in a very deep trance, their reports that the +H was indistinguishable from the actual field were not borne out by the objective measures of illusory effect. It must be recognized, however, that the study measured only a limited aspect of hypnotic hallucinations.

No definite conclusions can be reached from the study as to whether perception was altered by the +H. *S*s in a trance of a depth often accepted as sufficient for research in deep trance behavior did not show measurable figure distortion accompanying +Hs of the fields, even when they reported that the +H and actual field were indistinguishable.

There is suggestive evidence, however, that the positive hallucination of a field for *S*s who attained very deep trance did cause figure distortion. This trend was manifest in both measures of the +H (illusions A and C); it is magnified when the results for a deviant *S* are excluded.

Perhaps the simplest explanation for the trend for +H would attribute the results to chance. The near significant results and the inordinate effects of a deviant *S* are, of course, arguments against that explanation. A further bit of evidence to the contrary comes from the fact that three of the four hypnotic *S*s who showed any +H tendency showed it on all three figures of Illusion A and on Illusion C. The remaining two *S*s showed no signs whatsoever of the +H; in poststudy questioning they admitted having had incomplete hallucinations. A more random occurrence of correct responses would be expected in results due to chance.

Also supporting the idea that valid +Hs were obtained is the fact that among all the hypnotic *S*s, the four who showed +H effects had complete spontaneous amnesia, while all other *S*s had no such amnesia. The implication is that inadequate trance depth is responsible for the inconclusive results. Spontaneous amnesia, thorough training for a deep trance, and behavior that is completely consistent with a deep trance, would seem to be the minimal criteria for adequate trance depth.

Among the extraneous variables that might have been influential, major consideration should be given to the possibility that the *S*s inferred the correct responses from past knowledge or from cues in the experimental situation. Hypnotic *S*s are reputed to be particularly adept at using such information in faking responses. Although this study was designed specifically to prevent such an occurrence, it is not inconceivable that *S*s under hypnosis could recall some obscure childhood experience with the illusions or some such event that would allow them to fake their responses. This type of explanation does not, however, account for the fact that there was no evidence whatsoever for a valid −H while there was suggestive evidence for the +H. Any past knowledge or cues from the experiment should have been equally applicable under both conditions.

It is of considerable importance that the control and hypnotic groups were not made comparable prior to the initiation of experimental procedures. In selecting hypnotic *S*s, rigid criteria of trance depth were applied and the *S*s used in the study were screened from a very large number of potential *S*s. No such intensive screening could be used in selecting control *S*s. In addition to the fact that it is extremely difficult to find *S*s capable of very deep trance, it was felt that had controls been selected and trained to reach a deep trance as were the hypnotic groups, then one could not be certain that they would not enter a trance, perhaps inadvertently, under conditions of waking suggestion, thereby nullifying their function as a control group. The control group was selected solely for comparability in age and education level. It is quite possible that the groups could have differed in some relevant but uncontrolled way. In particular, the highly selective procedure used in obtaining hypnotic *S*s may have selected *S*s with unusually vivid visual imagery. Such a capacity might make it possible for them to imagine the illusion field with sufficient vividness to perceive the correct figure distortion whether they were hypnotized or not. This conjecture is, of course, contrary to the reports of very deeply hypnotized *S*s that in hyponosis, their hallucinations are indistinguishable from reality and when not hypnotized they only have fleeting mental images.

SUMMARY

The validity of hypnotic visual hallucinations was investigated by the use of three optical illusions, in which a field of lines distorts a superimposed figure. If hallucinations actually involve perceptions similar to those aroused by objective stimulation, then a figure superimposed on the hallucinated field should appear properly distorted and the distortion of a figure in an actual field should be removed if the absence of the field is hallucinated.

Six deeply hypnotized *S*s and six very deeply hypnotized *S*s reported perceptions of figures projected on a screen during both hallucinated and actual presence and absence of the field on the screen. Six nonhypnotized *S*s performed in the same conditions except that they imagined the presence and absence of the field and guessed the way figures would look if the suggested stimulus conditions actually existed.

1. Conclusive evidence was obtained that the negative hallucination did not measurably lessen the distorting effect of the field.

2. Conclusive evidence was obtained that for *S*s who satisfy only minimal criteria for hallucinations and deep trance, the positive hallucination of the field does not cause measurable figure distortion.

3. Suggestive evidence was found that for *S*s who are in a very deep trance, in which all behavior is thoroughly consistent with a deep trance and all visual hallucinations are produced with ease, the positive hallucination of a field does cause figure distortion, but to a degree less than that caused by objective field stimulation.

REFERENCES

BENUSSI, V. Zur experimentelin Grundlegung hypnosuggestiver Methoden psychischer Analyse. *Psychol. Forsch.*, 1927, **9**, 197–274.

BERLINER, A., & BERLINER, S. The distortion of straight and curved lines in geometrical fields. *Amer. J. Psychol.*, 1948, **61**, 153–166.

ERICKSON, M. H. A study of clinical and experimental findings on hypnotic deafness: II. Experimental findings with a conditioned reflex technique. *J. gen. Psychol.*, 1938, **19**, 151–167. (a)

ERICKSON, M. H., & ERICKSON, E. M. The hypnotic induction of hallucinatory color vision followed by pseudo-images. *J. exp. Psychol.*, 1938, **22**, 581–588. (b)

HENRIKSSON, M. The effect on an established percept of a perceptual process beyond awareness. *Acta Psychol.*, 1955, **11**, 346–355.

HOFMANN, B., & BIELSCHAWSKY, A. Über die Einstellund der scheinbaren Horizentalen und Vertikalen bei Betrachtung eines von schrägen Konturen erfülten Gesichtsfelds. *Arch. ges. Physiol.*, 1909, **126**, 453–475.

MOTOKAWA, K. Field of retinal induction and optical illusion. *J. Neurophysiol.*, 1950, **13**, 413–326.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer., 1953.

(Received April 29, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 47-51

PERSONALITY SIMILARITY AND FRIENDSHIP[1]

CARROLL E. IZARD[2]
Vanderbilt University

WHY DO Person A and Person B find each other attractive and each other's company a satisfying experience? Two antithetical answers have been formulated to this question in recent years. Winch (1958) argued that the principle of complementarity explains attraction between marriage partners: successful marriages are made up of people whose needs and characteristics complement each other. Newcomb (1956), working with college males who were brought together as strangers and shared common quarters for several months, pointed to perceived similarity as the most significant factor in interpersonal attraction: Person A is attracted to Person B if A perceives B as like A. Newcomb's similarity thesis was given some support by Kelly (1955), who, like Winch, used married couples as *S*s. Several other studies have found real or assumed similarity to be a significant factor in interpersonal relations (Fiedler, 1958; Rosenfeld & Jackson, 1959); however, Morton (1959)[3] reviewed the literature concerned specifically with personality similarity and friendship and found contradictory conclusions.

The principal thesis underlying the present study is a variant of Newcomb's similarity principle but differs in its emphasis on similarity of affective personality factors and on actual rather than perceived similarity. The affective component in interpersonal attraction has come to seem more important with the recent work of Tagiuri (1958) and Schutz (1958) on the significance of affective rapport—"mutuality," "congruency," "compatibility"—in interpersonal relations and group productivity.

These developments support the basic postulate of the present work, that interpersonal positive affect—the expression of favorable feeling, self-involving interest, acceptance, or esteem in relation to another person—is a significant determinant of individual and interpersonal behavior. A further underlying assumption is that personality similarity facilitates the expression of interpersonal positive affect.

The aim of this study was to throw some light on the general problem of interpersonal attraction by studying personality (affective) variables in relation to friendship. The specific hypotheses were (*a*) mutual friends have similar personality profiles, and (*b*) mutual friends have significant positive correlations on some of the separate personality characteristics that make up the profile.

METHOD

Over 200 students from a high school and a private college, both of which draw largely from the white middle and upper classes, were asked to list their closest personal friends in rank order. They were then given Edwards' (1954) Personal Preference Schedule (PPS), a forced choice personality inventory that measures 15 manifest needs originally defined by Murray (1938). Various PPS scales have been found significantly related to other personality tests (Silverman, 1957) and to relevant criteria derived from actual behavioral situations (Bernardin & Jessor, 1957; Gisvold, 1958; Izard, in press).

The sociometric data made possible the selection of 30 pairs who chose each other as best friends. In addition, a control group of 60 *S*s from an entering college class were selected and paired at random, with the restriction that there be the same number of male and female pairs as in the 30 pairs of friends.

An analysis of variance technique, described in detail in connection with presenting the results, was used to test the hypothesis that mutual friends have similar personality profiles. This method made it possible to test for profile similarity between pairs of friends and between random pairs, and to compare the group of friend pairs to the group of random pairs with respect to similarity. The technique gives considerably more information and tests of significance based on more degrees of freedom than would the more conventional

---

[1] This paper was presented at the 1959 meeting of the APA.

[2] The author is indebted to C. I. Bliss, Yale University and Connecticut Agricultural Experiment Station, for invaluable assistance in developing the analysis of variance procedure for comparing profiles; and to Dudley Peeler, Edna Johnson, David Catron, and Billie Farmer for assistance in obtaining and analyzing the data for this study.

[3] Gulliksen was principal investigator under Office of Naval Research (Contract Nonr 1858(15)), and National Science Foundation Grant G-642, Princeton: Princeton University and Educational Testing Service, 1959.

TABLE 1
ANALYSIS OF VARIANCE OF PPS SCORES OF FRIEND AND RANDOM PAIRS

Source	*df*	Sum of Squares	*MS*	*F*	*P*
Between Variables	14	$\Sigma V_t^2/N - T^2/N_o = A$	805.338	47.803	.001
Variables × Groups	14	$(\Sigma v_{t1}^2 + \Sigma v_{t2}^2)/N_g - C - A$	16.847	<1.000	
Variables × Pairs (Friends)	406	$\Sigma y_{t1}^2/N_p - \Sigma v_{t1}^2/N_g$	23.143	1.529	.01
Variables × Pairs (Random)	406	$\Sigma y_{t2}^2/N_p - \Sigma v_{t2}^2/N_g$	22.811	1.226	.05
Variables × *S*s (Friends)	420	$\Sigma y_1^2 - \Sigma y_{t1}^2/N_p$	15.138	1.229	.05
Variables × *S*s (Random)	420	$\Sigma y_1^2 - \Sigma y_{t2}^2/N_p$	18.599		
C (Correction)	1	T^2/N_o			

series of profile correlations or D measures. An r or D is useful as an index of similarity for a single pair, but neither is as powerful as the analysis of variance technique for studying a series of pairs and for comparing one such series with another with respect to profile similarity.

The hypothesis that mutual friends have significant correlations on some of the individual personality variables was tested by computing intraclass correlation (r_i) for each of the 15 personality characteristics. Corresponding correlations were computed for the *S*s paired randomly. The significance of an r_i was determined by Fisher's (1948) method: transform r_i to z, compute corrected z or $z_c = z + \frac{1}{2n' - 1}$, and then compare z_c with $s_{z_c} = \frac{1}{n' - \frac{3}{2}}$. The acceptable level of significance was set at .025, since testing the hypothesis required the computation of 30 r_i's.

RESULTS

The hypothesis that mutual friends have similar personality profiles was supported by the analysis of variance for comparing profiles. For this analysis, profiles were entered in adjacent rows of the data sheet. Since entries in a given row were not independent and summed to a constant, the terms usually based on row totals drop out. To compute the terms in the analysis, let y_1 and y_2 represent the scores of friends and random *S*s respectively on a given personality variable, y_{t1} and y_{t2} the respective totals for a friend and random pair on a variable, v_{t1} and v_{t2} the variable total for the friend group and random group respectively, V_t the total for both groups on a variable, T the grand total for all observations, N_p the number in a unit (pair), N_g the number in a group, N the number of *S*s, and N_o the total number of observations. The analysis of variance together with the computational formulas for the sums of squares is presented in Table 1.

The Between Variables term simply represents the variance associated with the average profile for the 120 *S*s. This term is of no consequence in the present study. It simply indicates that on the average the *S*s responded significantly differently to the 15 PPS scales. It was expected to be quite large in comparison with its error—the Variables × Groups term.

The Variables × Groups term is similar to a simple Between Groups term in a single variable analysis of variance. The principal difference is that here each group is measured on 15 separate variables. Hence, the variance due to variables can be isolated (Between Variables term) and the groups compared in terms of their differential interaction to the 15 variables (Variables × Groups term). The same logic applies to the other interaction terms in the analysis. The Variables × Groups term was tested against the pooled variance of the two Variables × Pairs terms. The resulting F was less than 1.000. This indicated that the average profile for the friend group did not differ significantly from that of the random group.

The test for profile similarity of friends

$$F = \frac{\text{Variables} \times \text{Pairs (Friends)}}{\text{Variables} \times S\text{s (Friends)}}$$

can be interpreted as a between variance–within variance ratio. The variance within pairs of friends (Variables + *S*s) was significantly smaller ($P < .01$) than the variance between friends (Variable × Pairs). However, the corresponding F for the random *S*s was also significant, though less so ($P < .05$). It was considered that this significant F for similarity of random pairs was possibly due to chance fluctuations in random sampling. However, two additional samples of random pairs from the same population had significant profile correlations. While further research is needed to determine the meaning of this finding, it makes it necessary to test whether friend pairs are significantly more alike than *S*s paired at random.

In order to compare the relative similarity of friend and random pairs, it was necessary to check whether the two groups were homogeneous with respect to overall variability. This test

$$F = \frac{\text{Variables} \times \text{Pairs (Random)} + \text{Variables} \times Ss \text{ (Random)}}{\text{Variables} \times \text{Pairs (Friends)} + \text{Variables} \times Ss \text{ (Friends)}}$$

($P > .025$), indicated that the profiles of the 60 *S*s in the friend group did not differ on the average from those of the 60 *S*s in the random group. The comparison of the within variances of the two groups of pairs

$$F = \frac{\text{Variables} \times Ss \text{ (Random)}}{\text{Variables} \times Ss \text{ (Friends)}}$$

($P < .05$), showed that mutual friends considering a pair as a unit, were significantly more similar than *S*s paired at random.

Table 2 presents the intraclass correlations testing the hypothesis that mutual friends have significant positive correlations on some of the personality characteristics that make up the profile. As shown in the table, no r_i was significant for the *S*s randomly paired. For mutual friends, r_i was significant at the .025 level for Exhibition, at approximately the .01 level for Deference, and beyond the .01 level for Endurance.

The results from both the analysis of variance and intraclass correlation are evidence for similarity and contrary to what would be expected from complementarity; namely, that the scores of friend pairs on a particular variable should correlate negatively. However, they do not completely rule out the possibility of another less well defined type of complementarity, in which a given personality characteristic (e.g., Dominance) complements another (e.g., Abasement). To test for the presence of such complementarity across characteristics, it is necessary to identify characteristics that are psychologically complementary. There is little theoretical or empirical basis for making these pairings. The most appropriate empirical basis available for these pairings is the magnitude of the negative intercorrelations of the 15 variables in the standardization group. These negative correlations might be interpreted as evidence of within person complementarity. Primarily on this basis, the seven most relevant of the 105 possible combinations were tested. The intraclass correlation is again appropriate, since there is no logical basis for assigning *S*s to *x* and *y* columns for an ordinary product moment correlation (Fisher, 1948). The estimate of correlation for a given pair of characteristics would be obtained by pooling the two r_is for that combination, and the acceptable level of *P* would be somewhat less than .01. The 14 r_is were essentially zero, the largest having a sign in the opposite direction to that predicted by the complementarity principle.

There is a possibility that the social desirability of characteristics may be a factor in the friendship pairings, although the fact that Edwards employed the forced choice technique to eliminate social desirability at the item level makes this unlikely. To check on the possibility, the 98 students in a general psychology course were asked to rank the 15 characteristics in terms of social desirability. The final ranks based on the *S*s' individual

TABLE 2

INTRACLASS CORRELATIONS FOR FRIENDS AND FOR RANDOM *S*s

PPS	Intraclass Correlations	
Variable	Friends	Random *S*s
Achievement	.200	.031
Deference	.435*	.347
Order	.279	−.157
Exhibition	.386*	.179
Autonomy	−.001	.060
Affiliation	.022	.138
Intraception	.113	−.183
Succorance	−.082	−.077
Dominance	.112	.201
Abasement	.168	.296
Nurturance	.100	.223
Change	.253	.212
Endurance	.473*	−.248
Heterosexuality	.116	.193
Aggression	.167	−.049

* $P < .025$.

rankings were normalized by means of Fisher's rankit transformation (Fisher & Yates, 1953). Since the rankits for the 49 men correlated .984 with the corresponding values for the 49 women, final rankits for social desirability were based on all 98 *S*s. The rankits for social desirability were correlated with rankits derived from the rank order of the characteristics in terms of the magnitude of the r_is for the friend group. This correlation was only .002, indicating that on the average there is no relationship between the social desirability of a given personality characteristic and the degree to which friends are similar on that characteristic. Further, it is interesting to note that the profile based on this index of social desirability did not correlate significantly with the obtained average profile for a large sample of college males or with the average profile for a large sample of college females.

DISCUSSION

There is some support for the importance of perceived similarity, assumed similarity, and actual similarity in interpersonal attraction and friendship. Newcomb (1956) found that perceived similarity was a key factor in interpersonal attraction, and Fiedler (1958) reported that assumed similarity was significantly related to leader attitudes and team effectiveness. Rosenfeld and Jackson (1959) showed that similarity on three scales from the Guilford-Zimmerman Temperament Survey (Sociability–Unsociability, Ascendance–Submissiveness, Security–Insecurity) was significantly related to friendship ratings of female office employees. The present study showed that mutual friends, people who are attracted to one another over an extended period of time, have personality profiles as measured by the 15 scale PPS that are significantly more similar than those of *S*s paired at random and that they have significant positive correlations on some of the separate personality characteristics.

The question arises as to why personality similarity is a factor in friendship or interpersonal attraction and as to how this factor operates in interpersonal dynamics. The role of personality similarity may be understood by viewing it in relation to the underlying hypotheses of the present study. It was postulated that interpersonal positive affect is an important determinant of interpersonal behavior and that personality similarity facilitates the expression of positive affect. It seems reasonable that the expression of positive affect—favorable feeling, self-involving interest, acceptance, esteem—enhances interpersonal attraction. But how does personality similarity facilitate positive affect? If Person A and Person B have similar personality profiles, they have vested affect or affect needs in common. More dynamically, they have similar ways of expressing and receiving affect, and this actual similarity of affective characteristics might reasonably be expected to facilitate mutually satisfying interactions and experiences. For one thing, the same individuals, groups and interpersonal situations that evoke positive affect in one member of an affectively similar pair would tend to evoke positive affect in the other. Actual personality similarity should also increase the accuracy of interpersonal perception and communication, particularly the perception and communication of attitudes, values and other affective aspects of experience. This is almost tantamount to saying that *S*s can perceive and communicate that which is similar to their own experience more readily than the dissimilar, the familiar more readily than the unfamiliar. This formulation is in keeping with Newcomb's (1956) conception that accurate interpersonal communication and perception of similarity are interdependent rewarding conditions. Insofar as a person sees his self-involving, affect laden desires and actions confirmed by the desires and actions of another, his perceptions of self and others receives confirmation. This line of thinking is also consonant with the concept of congruency—"the tendency, correctly based on experience, to perceive a person's feelings for us as congruent with our feelings for him"—the factor which Tagiuri (1958, p. 321) has considered the most significant determinant of social preference. In fine, personality similarity, or similar affect needs and similar ways of expressing and receiving affect, should facilitate interpersonal positive affect and make for effective communication, compatibility and cohesiveness in the two person group.

SUMMARY

The present study was designed to test the hypotheses that mutual friends (*a*) have similar personality profiles, and (*b*) significant positive correlations on some of the separate personality characteristics that make up the profile. Mutual "best" friends were determined by a sociometric procedure, and personality characteristics were measured by the 15 scale PPS.

An analysis of variance technique for comparing personality profiles showed that pairs of friends were significantly more similar than pairs established by random assignment. Hypothesis *b* was supported by significant intraclass correlations among friends on Exhibition, Deference, and Endurance and no significant correlations among the random *S*s. These results tend to rule out the possibility of within characteristics complementarity, and an additional analysis yielded evidence against across characteristics complementarity. Further analysis showed that the social desirability of the personality characteristics was not significantly related to the degree to which friends are alike on the characteristics or to the degree to which the characteristics are actually present in the population sampled.

Accepting friendship as a criterion of attraction between people and the PPS as a measure of affective characteristics, the present findings may be restated as follows: personality similarity or similarity of affect needs and of ways of expressing and receiving affect is a significant factor in interpersonal attraction. These findings were discussed in relation to Newcomb's similarity principle and Tagiuri's concept of congruency. Personality similarity was interpreted as essentially a facilitator of interpersonal positive affect, the latter being postulated as a key determinant of attraction as well as other aspects of interpersonal behavior.

REFERENCES

BERNARDIN, A. C., & JESSOR, R. A construct validation of the Edwards Personal Preference Schedule with respect to dependency. *J. consult. Psychol.*, 1957, **21**, 63–67.

EDWARDS, A. L. *Manual, Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

FIEDLER, F. E. *Leader attitudes and group effectiveness.* Urbana: Univer. Illinois Press, 1958.

FISHER, R. A. *Statistical methods for research workers.* London: Oliver, Boyd, 1948.

FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural and medical research.* New York: Hafner, 1953.

GISVOLD, D. A validity study of the Autonomy and Deference subscales of the EPPS. *J. consult. Psychol.*, 1958, **22**, 445–447.

IZARD, C. E. Personality characteristics associated with resistance to change. *J. consult. Psychol.*, in press.

KELLY, E. L. Consistency of the adult personality. *Amer. Psychologist*, 1955, **10**, 659–681.

MORTON, A. S. Similarity as a determinant of friendship: A multidimensional study. *USN tech. Rep.*, 1959, Nonr 1858 (15), Princeton.

MURRAY, H. A. *Explorations in personality.* A clinical and experimental study of fifty men of college age. New York: Oxford, 1938.

NEWCOMB, T. M. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–586.

ROSENFELD, H., & JACKSON, J. Effect of similarity of personalities on interpersonal attraction. *Amer. Psychologist*, 1959, **14**, 366–367 (Abstract).

SCHUTZ, W. C. *FIRO: A three dimensional theory of interpersonal behavior.* New York: Rinehart, 1958.

SILVERMAN, R. E. The Edwards Personal Preference Schedule and social desirability. *J. consult. Psychol.*, 1957, **21**, 402–404.

TAGIURI, R. Social preference and its perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford Univer. Press, 1958.

WINCH, R. F. *Mate-selection: A study of complementary needs.* New York: Harper, 1958.

(Received June 8, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 52-66

CORTICAL INHIBITION AND EXTRAVERSION-INTROVERSION

WESLEY C. BECKER

University of Illinois[1]

THE RESULTS of several related experiments have led Eysenck to hypothesize that cortical inhibition is one of the bases for individual differences in extraversion-introversion. The extravert is assumed to generate more cortical inhibition faster and to dissipate it more slowly than the introvert. According to Eysenck (1955), cortical inhibition builds up in any neural structure mediating a stimulus-response connection as a result of excitation of that structure. Eysenck has further assumed that Köhler and Wallach's (1944) satiation concept and Hull's (1943) reactive inhibition concept can be partially subsumed under the concept of cortical inhibition. Satiation effects are usually inferred by measuring the effect of stimulation on subsequent perceptual responses. Reactive inhibition is usually inferred by measuring the effect of making a response on subsequent motor responses. Köhler and Wallach considered their satiation effects to be operating within the central nervous system, while Hull referred only to "physical structures" as the locus of action for reactive inhibition, leaving open the question of the effects as being neural or muscular or both.

Eysenck has at times used the terms reactive inhibition, cortical inhibition, and satiation interchangeably, thus neglecting the historical differences in meanings and fostering some confusion. In one paper (1955), cursory reading would lead one to believe that Eysenck has fully embraced Hull's reactive inhibition concept as the primary construct in his theory, but this is not actually the case, since he modified the definition of reactive inhibition to rule out muscular effects. The resulting construct is more appropriately called *reactive cortical inhibition*. This concept, as used by Eysenck, is assumed to cover what is common to the operations involved in defining satiation and reactive inhibition, namely, the effect on subsequent behavior of the repeated excitation of a neural structure mediating a stimulus-response connection. However, this concept is not assumed to cover the additional operations (such as magnitude being dependent on the amount of work involved) which might be used to define reactive inhibition or satiation.

Eysenck has published two studies which have supported the hypothesis that reactive cortical inhibition is one of the bases for individual differences in extraversion-introversion. In the first study (1955), he found significant differences between male hysterics (i.e., extroverted neurotics) and dysthymics (introverted neurotics) in the amount of aftereffect produced by kinesthetic stimulation on tactile judgments of size. The greater cortical inhibition effects produced in extraverts was believed to have led to decreased judgments of size following stimulation. In this study, Guilford's Rhathymia Scale was used in addition to psychiatric diagnosis to define extraversion-introversion. In a second study using normal male university students, Eysenck (1956) argued that extraverts with greater reactive cortical inhibition show more reminiscence on a pursuit rotor task than introverts. In this experiment, the operations defining cortical inhibition paralleled those for reactive inhibition, whereas in the first experiment the operations paralleled those for satiation. His findings indicated a significant correlation of .29 between the Maudsley Personality Inventory Extraversion Scale and the first reminiscence score, but an insignificant correlation of .10 with a second reminiscence score. In a personal communication, Eysenck indicated that one exact replication and one replication using a different task have given similar significant results. On the other hand, a study using 90-sec. performance trials before rest (rather than five-minute trials) failed to show the expected relationships.

Eysenck has also referred to a conditioning study by Franks (1956) as supporting his hypothesis. Franks found that dysthymics conditioned more readily than hysterics, using the eyeblink reflex and simultaneously the GSR. Like Eysenck, Franks attributed his results to the greater cortical inhibition in the extrovert. But unlike Eysenck, he made it clear that he used the concept of cortical inhibition to describe a more or less enduring trait. Franks wrote, "hysterics are in a state of cortical inhibition in which dissociation phenomena predominate . . . dysthymics are in a state of cortical excitation. . ." (p. 149). There was no indication in Franks' paper that he regarded cortical inhibition as a function of the number of excitations of neural structures mediating stimulus-response connections. To the contrary, he appeared to hold that individuals vary in their basal levels of cortical inhibition and that this was sufficient to account for his results. In a subsequent paper reporting a repetition of his doctoral study on normal *S*s, Franks (1957) shifted his theoretical ground and adopted Eysenck's *reactive* inhibition explanation to account for his findings. As Rechtschaffen (1958) has pointed out, however, it is difficult to defend a *reactive* inhibition hypothesis with eyelid conditioning data because the intertrial interval is not favorable for the accumulation of reactive inhibition. Even if the intertrial interval were favorable for such accumulation, Franks' acquisition curves fail to reveal any systematic tendency to diverge as should be expected if reactive inhibition were accumulating. In Franks' studies, differences were present on the first few trials and maintained thereafter.

---

[1] The author wishes to express his gratitude to the University of Illinois Graduate Research Board and to the United States Public Health Service (Project No. M-1713) for their financial support of this project.

A distinction between *basal* and *reactive* cortical inhibition has been proposed previously by Klein and Krech (1952). However, they chose to talk in terms of concepts which are operationally the inverse of inhibition. Their distinction is stated in terms of basal and reactive cortical conductivity rather than inhibition. It is interesting to note that Klein and Krech assumed a high positive correlation between individual differences in basal cortical conductivity and reactive cortical conductivity. If such a correlation could be demonstrated, then an easier way to integrate Franks' results with those of Eysenck would be available.

From this review, it is apparent that three operationally distinct concepts have been associated with the concept of cortical inhibition in one way or another, satiation, reactive inhibition, and basal cortical inhibition. In designing the experiment to be reported here, measurement procedures were included which would cover each of the three inhibition concepts so that the appropriateness of treating all three kinds of measures under a general cortical inhibition concept could be examined.

The idea of relating personality differences to attributes of the nervous system is very enticing. If Eysenck's suggestive results could be expanded upon and repeated by others, a wide area of behavioral study would be opened to more systematic exploration with the promise of unheard of integrations between neurophysiology, learning theory, and personality theory. However, at the start, a number of questions dampen this hope. Rechtschaffen (1958) has rightfully raised a question about the statistical significance of the kinesthetic aftereffect study. Farber, Spence, and Bechtoldt (1957) have found individual differences in eyelid conditioning related to anxiety and *not* to extraversion-introversion. Meier (1956) has reported a negative relationship between kinesthetic figural aftereffects and reminiscence in motor learning. This finding raises the question of whether Eysenck is justified in subsuming satiation effects and reactive inhibition effects under the same construct. Finally, in evaluating Eysenck's results, one is faced with the ever recurring problem of whether findings applicable to selected groups are generalizable beyond these groups. These were some of the hopes, doubts, and confusions which prompted this study.

The main hypotheses to be examined are: (*a*) that individual differences in satiation effects and reactive inhibition effects are correlated, thus, justifying the more general concept of reactive cortical inhibition; (*b*) that individual differences in basal cortical inhibition are correlated with individual differences in reactive cortical inhibition, as assumed by Klein and Krech (1952); and (*c*) that individual differences in satiation, reactive inhibition, or basal cortical inhibition are related to individual differences in extraversion-introversion.

METHOD

Thirty-two male and 30 female students were selected at random from a pool of nearly 1000 introductory psychology students. The mean age of the *S*s was 20.2 with a *SD* of 2.0. Father's education averaged 11.8 grades with a sigma of 3.5. Because the experiment required eight hours per *S*, $1.00 was paid for each hour's time in excess of the five hours' time each *S* was required to serve as a part of his course requirements.

Assessment of Extraversion-Introversion

Each *S* was given Guilford's Temperament Schedules (13 factors) and Cattell's Revised 16 PF (Forms A and B). In his work, Eysenck has used two criteria in addition to the hysteric-dysthymic diagnosis. At times he has used Guilford's Rhathymia Scale (R) and at others the Maudsley Extraversion Scale (E), which is scored from Guilford's tests. For this study, Guilford's Rhathymia, Eysenck's Extraversion, and an extraversion score derived from the 16 PF by summing factors A, H, E, and F were considered critical for defining extraversion-introversion.

Assessment of Cortical Inhibition

A variety of experimental tests were designed or selected which would produce scores interpretable as measures of basal cortical inhibition, satiation, or reactive inhibition. Since one must stimulate a *S* to get a basal type measure, such measures would not be entirely free of reactive effects. On the other hand some reactive measures cannot be assumed to be free of basal differences. Usually it is only possible to say that the basal variance is probably larger than the reactive variance for such and such a variable, while the reactive variance is larger than the basal for another measure. Of the eight experimental tests used, two clearly measured reactive inhibition effects (pursuit rotor reminiscence and response alteration), three fitted the definition of a satiation effect (kinesthetic aftereffect, Archimedes spiral, and the Necker cube difference score), while three approximated basal type measures (GSR conditioning, aniseikonic lenses, and CFF). Some exceptions to these general classifications are noted later.

The tests were given in three sessions. The first two-hour testing session included GSR conditioning, pursuit rotor, aniseikonic lenses, and Necker cube tests. The second two-hour session included kinesthetic figural aftereffect, CFF, Archimedes spiral, and alternation behavior tests. The third one-hour session, which occurred one to two weeks later, involved retesting on GSR, pursuit rotor, Necker cube, Archimedes spiral, and kinesthetic figural aftereffect.

In the interest of space, only a brief description of each test device is given below. The detailed testing procedures and rationales can be found in Becker (1959) or may be obtained from the author. However, information concerning reliability and construct validity of the measures will be discussed in detail so that an adequate interpretation of the findings is possible.

Measures of Basal Cortical Inhibition

GSR Conditioning. Because most differences in conditionability are apparent in the first few trials, the writer favors Franks' first interpretation of conditioning measures as reflecting basal differences in cortical inhibition. The method consisted of presenting twice by tape recorder a list of 104 words each separated by 5 sec. The word "repeat" was followed by shock on 12 of 20 trials during the first cycle and by a buzzer on 12 of 20 trials during the second cycle. On retest, the list was repeated for one cycle with shock as the US. A Hunter Model 300 GSR amplifier was used to record maximum resistance drop to each stimulus.

The most frequently used measure of conditioning is the number of test trials on which a conditioned response occurs. However, it was found that this kind of measure was greatly influenced by the general reactivity of the *S* as indicated by the magnitude of his response to the US. A measure which was found to be independent of this influence was the number of trials on which the CR was one-half or greater than the preceding UR. This measure, in addition, compensates for the typical adaptation of the UR and CR over trials and for artifact related to basal resistance level and shock level. By this definition, conditioning is considered stronger the closer the CR approximates the UR. Two generalization measures were also available. One based on temporal generalization and the other based on generalization in terms of sound similarity. These measures were combined to obtain a more reliable generalization measure. The generalization measure was also corrected for reactivity by dividing the average resistance drop to the generalization stimuli by the average resistance drop to the US.[2]

The correlations between conditioning measures are presented in Table 1. These intercorrelations can be interpreted as lower bounds of the retest reliability coefficient. They are lower-bound coefficients because they include errors of measurement plus any changes in factor composition of the scores over successive blocks of trials. The combined first session score has a reliability of .80 when the Spearman-Brown estimation formula is applied and correlates .71 with the conditioning measure derived from the second session.

While the usual learning curve type of construct validity evidence is not possible with GSR data because of the complicating adaptation effects, it is possible to compare the resistance drop to the CS and to the generalization words with the reaction to a control word "constant" (which appeared 11 times per cycle but never adjacent to the US) and to the other filler words which appeared but once on each cycle. Table 2 presents such an analysis. In each case the reported mean reaction is based on a *ratio* of a particular set of responses to the *S*s' response to the US. These measures are therefore analogous to the conditioning measure discussed above, based on the number of trials on which the CS is one-half or greater than the preceding US.

From Table 2 it is apparent that the reaction to the conditioned stimulus is about seven times greater than

---

[2] This ratio type measure was found to correlate .93 with the measure based on the number of trials on which CR was ½ or more than the preceding UR.

TABLE 1

CORRELATION AMONG CONDITIONING MEASURES

Variables	1b	2	Generalization
1a Shock series, first session	.67	.66	.51
1b Buzzer series, first session	—	.64	.58
2 Shock series, retest session		—	.58
1 First session combined		.71	.59

TABLE 2

EVIDENCE FOR THE VALIDITY OF THE CONDITIONING MEASURES

Stimulus	Resistance Drop to Indicated Stimulus Relative to Resistance Drop to US	
	Mean	Sigma
Conditioned Stimulus		
Repeat (test trials, first session)	46.6%	
Repeat (test trials, second session)	40.1%	
Average to repeat	43.3%**	22.65
Generalization Stimuli		
Crafty (first session)	13.2%	
Crafty (second session)	27.9%	
Receipt (first session)	11.2%	
Receipt (second session)	17.8%	
Average to crafty and receipt	17.6%**	14.66
Control Stimuli		
Constant (first session)	6.4%	
Constant (second session)	6.2%	
All other words (first session)	6.0%	
All other words (second session)	5.2%	
Average to control stimuli	5.9%	11.50

** Significance of difference from control stimuli mean is beyond the .001 level.

that to the control word *constant* and to all other words in the list. Similarly, the reaction to the generalization stimuli is about three times greater than that to the control stimuli. Both of these differences are highly significant (CR = 11.6 and 4.8 respectively) and are consistent with interpreting the proposed measures as indices of conditioning.

Aniseikonic lenses. Aniseikonic lenses have had their main use in the study of stereoscopic perception. The induced effect of the lenses is to make a table appear to tip up like a drawing board or to make a wall lean toward the observer. Becker (1954) has demonstrated marked individual differences in the induced effect and has related the perceived magnitude of distortion and the time required to see the distortion to inhibited and uninhibited personality characteristics. The hypothesis offered is that extroverts with higher cortical inhibition should be less bound to their past perceptual habits and therefore more likely to shift in the new situation. A clear differential prediction, however, between satiation and basal effects can not be made for it is possible to argue that either might lead to weaker perceptual habits in this case. As in the conditioning case, the author is inclined to favor a basal difference hypothesis.

The apparatus and procedures duplicate those used by Becker (1954). Two scores were computed for later analysis: (*a*) *time delay*, which is the average time taken to report an effect on trials three to six, and (*b*) *estimated tilt*, which is the average tilt reported on trials three to six. Becker (1954) has previously reported a stability coefficient of .70 over a three week period for the time delay score. Using an analysis of variance approach, the reliability (consistency) of both scores was found to be .95.

Critical fusion frequency (*CFF*). Flicker fusion measures have often been considered relevant to the evaluation of the efficiency of cortical functioning (Landis, 1951). It is possible to predict that a higher basal cortical inhibition level would act to lower the perception of discrete flashes (lower CFF). However, since the organism is being stimulated over a period of time, satiation effects could also be operating. In an attempt to isolate basal and satiation effects, two kinds of flicker measures were taken, one freer of satiation effects than the other. The first was based on the usual adjustment method, wherein the rate of flicker was varied until fusion or flicker was reported. This score, based on the average of eight alternating up-down trials taken after 10-min. practice and suitable dark adaptation, will be referred to as CFF *up-down*. The second score was based on one-second exposures of the stimulus at fixed speeds while the subject merely reported "flicker" or "no flicker." This score will be referred to as CFF *median*, since it was based on the median of the interval bounded by four successive reports of "flicker" and "no flicker." It was assumed that the first measure would reflect basal inhibition and satiation effects, while the second measure would reflect basal effects primarily. The difference between the two measures was assumed to reflect a satiation effect. The apparatus consisted of an episcotister model after that used by McNemar (1941), except that a Compur shutter was mounted in front of the viewing glass to permit control of exposure time.

The correlation between the CFF up-down score and the CFF median score was .58. This should not be too high if the difference score is to have any value. Since previous work (McNemar, 1951) has indicated within sessions reliabilities above .95 for CFF judgments after practice, no further check of this was made. The construct validity evidence for these measures is at best suggestive. The CFF median procedure produced an average CFF which was 2.3 cps higher than that found by the up-down procedure. This difference is statistically significant (CR = 4.1) using the appropriate test for the difference between correlated means and is consistent with the satiation hypothesis offered earlier.

Measures of Satiation

Archimedes spiral. In a personal communication, Eysenck has suggested that satiation effects should *de-*

TABLE 3

CORRELATIONS, MEANS, AND SIGMAS FOR ARCHIMEDES SPIRAL MEASURES

Variables	Correlations			*M*	σ
	1b	1c	2		
1a Two-min. satiation	.60	.72	.28	2.29	3.58
1b One-min. satiation	—	.48	.38	1.90	2.71
1c Three-min. satiation		—	.66	2.68	4.12
2 Two-min. satiation, retest			—	2.82	3.94
1 Average of 1a, 1b, and 1c			.53		

TABLE 4

INTERCORRELATIONS, MEANS, AND SIGMAS FOR NECKER CUBE DIFFERENCE SCORES

Variables	Correlations			*M*	σ
	1a	1b	2		
1a One-min. trials	—	−.10	.23	.16	3.75
1b Three-min. trials		—	.38	1.69*	5.15
2 Two-min. trial, retest			—	−.26	2.70

* Significant at the .01 level.

crease the movement reversal aftereffect produced by looking at a rotating spiral. To measure the *length* of afterimage the writer modified the Archimedes spiral test from its recent use by Price and Deabler (1955). A two-revolution spiral was cut on a 1½ in. cylinder of plexiglass and attached to a phonograph motor within a light box. Lights in the box were controlled by a timing circuit which made it possible to have the lights continuously on or off, or to have them cycle on for one sec., off for three sec. *S* viewed the spiral with monocular tube vision. After appropriate instructions and a practice trial, *S* was given satiation trials of two-min., one-min., and three-min. duration. The length of afterimage score was based on the number of light flashes for which contraction of the spiral was reported before four cumulative failures to report contraction. An additional two-min. satiation trial was given during the retest period. In view of their moderately high intercorrelations, the first session measures were averaged. Table 3 presents the intercorrelations of the spiral afterimage measures, their means, and sigmas. Using the alpha coefficient, the within sessions reliability of the averaged first session measure was found to be .82. The between sessions reliability is tolerable only for the averaged measure and for the three-min. satiation measure. The fact that the mean length of afterimage shows systematic increase with length of satiation is suggestive evidence for construct validity. The mean difference between the one-min. and three-min. satiation periods is significant at the .05 level (CR = 1.68) using a one-tailed test.

Necker cube. McDougall, Cattell, and Eysenck have all tried to relate Necker cube measures to extraversion with little success. Eysenck believes the extravert's tendency to look away from the fixation point is the basis for the lack of consistent findings. The present writer is willing to grant that rate of oscillation of reversible figures is probably a function of a neural speed variable (Cattell, 1957) which is independent of cortical inhibition. However, it is possible to predict from the cortical inhibition hypothesis that extraverts should show a greater slowing down of reversible perspective over a period of continuous stimulation as a function of satiation. To guard against the extravert's alleged tendency to "leave the field," the stimulus was presented in a light box, using monocular tube vision. A typical Necker cube with one side shaded was presented to *S* for two one-min. trials and two three-min. trials. One two-min. trial was given on retest. A constant speed kymograph operated by a button was used by *S* to report reversals. The correlations among the difference scores and the means and sigmas are presented in Table 4. Since *S* was given two one-min. trials and two three-min. trials it was possible to compute reliabilities separately for each of these conditions. The reliability of the average of the two one-min. trial difference scores was found to be .17, and that for the three-min. trials, .42. On the basis of the mean difference data, it is apparent that only the three-min. trials produced a significant decrement as would be expected if a satiation effect were being measured. In view of this and the reliability findings, only the three-min. trial measure was retained for further analysis.

Kinesthetic figural aftereffect. This device is presumably a direct measure of satiation effects. *S* first makes a series of width judgments using only kinesthetic cues, and then, following stimulation of the fingers, the judgments are repeated. Eysenck (1955) gives details of rationale and apparatus. In the present experiment, eight baseline readings were made, alternating up and down from random starting points. Then four test judgments were made following satiation periods of 30, 60, and 120 sec. Four more test judgments were taken after 10 min. rest. A 60 sec. satiation period was used on retest after establishing baseline once more. Rate of satiation was controlled by having *S* move his hand in rhythm with a metronome operating at 126 beats per min. The satiation measure is that used by Eysenck (percentage decrement) and is found by subtracting the average postsatiation judgment from the average baseline judgment and dividing by the average baseline judgment.

Eysenck (1955) has reported reliabilities based on an analysis of variance estimate of *r* ranging from .78 to .94 for decrement scores. However, these are undoubtedly spurious. When the same baseline scores are used to compute two decrement scores which are then intercorrelated, a variety of index correlation is produced which has spurious properties. It becomes possible to obtain a correlation of .50 between two decrement scores when in fact there is no correlation between the test judgments or baseline judgments upon which the decrement scores are based. With the means and variances in the present study, the intercorrelations of the four decrement scores from the first session would be about .40 even if there were no correlations (no reliability) between the test judgments and basal judgments involved.

TABLE 5

Kinesthetic Figural Aftereffects

Variables	Correlations, Basal Partialed Out		
	1b	1c	1d
A. Intertrial Reliability Estimates, First Session			
1a Mean test judgment, 30″ trial	.56	.55	.27
1b Mean test judgment, 60″ trial	—	.70	.38
1c Mean test judgment, 120″ trial		—	.54
1d Mean test judgment, 10′ delay			—
Variables	**Correlations with Retest Decrement Score**	***M***	***σ***
B. Between Sessions Reliability, Means, and Sigmas			
1a Percentage decrement, 30″ trial	.05	8.83	10.8
1b Percentage decrement, 60″ trial	.21	8.97	12.9
1c Percentage decrement, 120″ trial	.10	11.26	14.7
1d Percentage decrement, 10′ delay	−.21	8.12	13.8
1 Mean percentage decrement, first session	.04	9.29	11.3
2 Percentage decrement, 60″ retest	—	5.96	11.8

In view of this, an attempt was made to gain an estimate of the within-session reliability by computing correlations between the various test judgments with baseline judgment partialed out. The results of this analysis are presented in Part A of Table 5. The pattern of the intercorrelations in Table 5 suggests that whatever the judgments are measuring is changing somewhat from trial to trial. However, there appears to be enough common variance to justify adding the four estimates to gain a single more reliable measure. The alpha coefficient, based on the correlations among the test judgments with basal level partialed out, gives a within sessions reliability estimate of .80 for the four averaged scores. Part B of Table 5 indicates that the retest reliability of the decrement scores is nil. Neither Eysenck nor Klein and Krech have reported retest reliability data. It would appear that a stable individual difference is not being measured. This finding raises a serious question about the adequacy of these figural aftereffect scores for testing Eysenck's hypothesis.

When the means are examined, it becomes apparent that satiation-like effects did occur and that they are of the same order of magnitude as those reported earlier by Klein and Krech (1952) and by Eysenck (1955).[3] These indications coupled with the finding of adequate within session reliability suggest that the present experimental procedures have replicated those of Eysenck as intended.

Measures of Reactive Inhibition

Pursuit Rotor Reminiscence. If, as Hull has argued (1943), reminiscence effects are as function of the dissipation of reactive inhibition (I_R), it follows from Eysenck's hypothesis that extraverts should show more reminiscence.[4] Using a standard pursuit rotor apparatus, each *S* was given up to 20 practice trials of 30 sec. duration, separated by 10 sec. rest. When a male reached a criterion of 14 sec. on target, or a female a criterion of 12 sec., for two consecutive trials, the practice period was ended. The average number of trials to criterion was 13.2 ($\sigma = 4.6$). Three males and five females were eliminated for failure to reach criterion. After a 5-min. rest, two 7-min. continuous performance trials were given separated by 5-min. rest periods. The final rest was followed by a 2-min. trial. During retesting, an additional reminiscence score was obtained using a 7-min. performance trial, 5-min. rest, and 2-min. performance. Reminiscence scores were based on the difference between the minute's performance preceding each 5-min. rest and the minute's performance following each 5-min. rest.[5] The correlations among the reminiscence scores are given in Table 6. Scores based on the average of the two first-session scores and the retest score were retained for later analysis. The consistency coefficient for the averaged first session score was found to be .69. The lower bound of the stability coefficient for this averaged score was .44. Significant reminiscence effects were obtained for all three measures, although there appears to be a systematic reduction in reminiscence effects over test trials.

---

[3] For 30″, 60″, and 120″ satiation periods and the 10 min. delay, Eysenck's means were 5.7, 8.4, 10.4, and 4.2; while those reported by Klein and Krech were approximately 6.0, 8.5, 12.0, and 9.0. In a study completed after this paper was written, Northrup and Becker (in preparation) have demonstrated that judgment artifacts can produce similar decrement effects when no satiation is given, the *S* merely being asked to make repeated judgments after rest periods corresponding to the satiation periods.

[4] Rechtschaffen's (1958) argument that the slower dissipation of reactive inhibition in extroverts leads to a conflicting prediction is not critical since Kimble and Horenstein (1948) have demonstrated that most I_R effects are dissipated within five min. Thus, if five min. rests are used, reminiscence scores should largely reflect I_R.

[5] Eysenck (1956) based his reminiscence score on the difference in performance on the last 30 sec. before rest and the first 10 sec. after rest. This scoring was based on the assumption that the warm-up effect after rest (Ammons, 1947) was due to the extinction of conditioned inhibition (${}_SI_R$) and therefore, taking a longer postrest interval would confound I_R and ${}_SI_R$. Finding no evidence to support this theory and finding it difficult to accept in the first place, a scoring procedure more consistent with Ammons' reasoning (1947) was adopted. Experience with the highly unstable 10 sec. measures in preliminary studies also influenced the decision toward a potentially more reliable measure.

TABLE 6

INTERCORRELATIONS, MEANS, AND SIGMAS FOR REMINISCENCE SCORES

Variables	Correlations		M	σ
	1b	2		
1a First session, first reminiscence score	.52	.41	14.8	6.7
1b First session, second reminiscence score	—	.37	12.1	6.2
2 Retest session reminiscence score		—	8.0	7.2
1 Average of 1a and 1b		.44		

Alternation Behavior in a DRT Setting. A classical application of the I_R concept has been in the interpretation of alternation behavior. In an attempt to measure individual differences in alternation behavior, a discrimination reaction time apparatus was built which permitted E to turn on one of three lights (right, middle, left) which S could then turn off with one of two levers (right, left). The middle light could be turned on so that only the right or left lever would turn it off. The right and left lights could be turned off by the right and left levers, respectively. This made it possible for E to get any combination or series of responses with the right and left hands and then to see which hand was preferred when the middle light was turned on. To prevent S from forming hypotheses about sequences, pseudo ESP instructions were given along with the idea that a random sequence was being used. The conspicuous electric timer kept S focused on quick responses. After giving suitable instructions, a prearranged program of lights was presented. The program consisted of five parts: (*a*) 10 trials of right and left lights for initial reaction time measurement, (*b*) a *random* series of 60 right, middle, and left lights, (*c*) 87 trials during which *runs* of right and left lights varying from one to five precede the middle light, (*d*) Part b repeated, and (*e*) Part a repeated. The response alternation score was the percentage of trials on which S responded to the middle light with an alternation response. These scores were computed separately for Parts b, c, and d. It was expected that Part c would give the highest percentage of alternations since more I_R would build up during a run in which the same response was made several times. Table 7 presents a confirmation of this expectation along with other relevant reliability and construct validity data. The findings in Table 7 suggest some construct validity, but a longer testing period is clearly needed to increase reliability to an acceptable level. For the subsequent analysis, the alternation scores from Parts b and d were averaged.

TABLE 7

CORRELATIONS, MEANS, AND SIGMAS FOR ALTERNATION RESPONSE MEASURES

Variables	Correlations		M	σ	CR
	B	C			
A. Percentage alternation responses Part b (random)	.25	.30	52.08	13.5	1.22
B. Percentage alternation responses Part c (runs)	—	.37	63.11	18.02	5.75**
C. Percentage alternation responses Part d (random)		—	54.90	13.3	2.92*
D. Average of Parts b and d	.38	—	53.5	10.8	2.55*

* Significantly different from chance beyond the .01 level, one-tailed test.

** Significantly different from chance beyond the .001 level, one-tailed test.

Summary of Status of Experimental Measures

Table 8 provides a resume of the reliability and internal validity evidence for the experimental measures retained for further analysis. It should be noted that a good proportion of the experimental measures showed mean variation with variation in stimulus conditions which were consistent with satiation or reactive inhibition theory (construct validity evidence). However, the reliabilities of many of the measures leave much to be desired. There will be need to refer back to this table in evaluating the results of later analyses.

RESULTS

The 17 experimental test measures, 18 experimental background measures, 3 sample background measures, 15 personality questionnaire variables, and 5 random variables were intercorrelated and submitted to a centroid factor analysis (fixed, unit communalities) and then rotated to simple structure using the varimax criterion (Kaiser, 1958).

In considering the results, the *r*s among the experimental measures will be examined first. These *r*s are presented in Table 9. The variables have been ordered to conform with the ordering of the summary given in Table 8, and the signs have been changed so that a positive *r* is consistent with the hypothesis of a general inhibition factor. Examination of Table 9 indicates that there is very little in common among the various experimental measures. Out of 136 possible *r*s, only 20 are significant in the predicted direction at the .05 level using a one-tailed test ($r = .21$). There are 6 significant *r*s in the direction opposite to prediction. If one considers only the *r*s which do not involve retesting with essentially the same measure, only 12 are significant in the expected direction.

The picture is not quite so discouraging when one considers only that submatrix of variables with fairly good reliability and

TABLE 8

SUMMARY OF STATUS OF EXPERIMENTAL MEASURES

	Within Session Reliability	Between Sessions Reliability	Construct Validity Evidence
A. Measures probably reflecting basal cortical inhibition			
1. GSR, conditioning (1)[g]	.80[a]	.71	good
2. GSR, conditioning (2)	.71[b]	.71	good
3. GSR, generalization (1, 2)	.59[b]	na[f]	good
4. Aniseikonic, time delay (1)	.95[a]	.70[c]	na
5. Aniseikonic, estimated tilt (1)	.95[a]	na	na
6. CFF, up-down series (1)	.99[d]	.70[d]	na
7. CFF, median, one sec. measures (1)	.58[b]	na	na
B. Measures of satiation effects			
1. CFF, difference score (1)	na	na	fair
2. Spiral afterimage (1)	.82[e]	.53	fair
3. Spiral afterimage (2)	.53[b]	.53	fair
4. Necker, difference score, three-min. trials (1)	.42[a]	.38[b]	fair
5. Kinesthetic aftereffect decrement (1)	.80[e]	.04	good
6. Kinesthetic aftereffect decrement (2)	na	.04	fair
C. Measures of reactive inhibition			
1. Pursuit rotor reminiscence (1)	.68[a]	.44	good
2. Pursuit rotor reminiscence (2)	.44[b]	.44	good
3. Response alternation, average Part b and d	.38[b]	na	poor
4. Response alternation, average Part c (1)	.38[b]	na	fair

[a] Spearman-Brown corrected.
[b] Lower-bound estimate.
[c] From Becker (1954).
[d] From McNemar (1951) using similar apparatus and procedure.
[e] Cronbach Alpha coefficient.
[f] na stands for "not available".
[g] (1) refers to first testing and (2) refers to retesting.

validity evidence. By dropping from the matrix the kinesthetic aftereffect measures (28, 31) which had zero retest reliability, the two difference score measures (38, 41), and the response alternation measure based on random series (42), it is found that 20 of the 66 *r*s in the remaining matrix are significant in the predicted direction. None is significant in the direction contrary to prediction. Furthermore only 6 out of 66 possible relationships show negative signs, whereas 42 are positive and greater than .10.

These *r*s suggest that there may be some significant relationship among the various measures of basal inhibition, satiation, and reactive inhibition, but the present measures are too unreliable to definitively establish such relationships. Examination of the *r*s in Table 9 suggests that the evidence is strongest for the existence of correlations among the basal inhibition measures, and that satiation and reactive inhibition measures while tending to correlate with the basal measures, do not tend to correlate with each other.

Factor Analysis of Variables

In order to save space the index of variables, the correlation matrix, the centroid factor matrix, and the varimax factor matrix for the 58-variable analysis have been deposited with ADI.[6] However, part of the varimax factor matrix is reported in Table 10, and the salient variables on most of the eight rotated factors are discussed in the subsequent text. Guilford and Cattell personality measures (except Rhathymia) which contribute to second-order anxiety or extraversion factors were omitted from the factor analysis to avoid inflation of loadings due to correlated errors.

The first problem in the interpretation of the factors is to decide which, if any, can be considered basal inhibition, satiation, or reactive inhibition factors. To aid in this decision, the critical variables were examined factor by factor to see to what degree the direction and size of loadings conformed to hypothesis.

It can be seen from Table 10 that Factor 3 shows consistently high loadings in the predicted direction on all of the basal inhibition measures and has loadings in the predicted direction (over .10) on 3 of the 6 satiation measures and on 3 of the 4 reactive inhibition measures. On the basis of sign alone, Factor 3 loads in the predicted direction 14 of the 17 variables. No other factor approaches Factor 3 in goodness of fit to the hypothesis of a

---

[6] Order Document No. 6315 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 9

CORRELATIONS AMONG EXPERIMENTAL MEASURES OF INHIBITION[a]

Variables	17	18	21	46	47	39	40	41^b	32	33	38^b	28^b	31^b	24	25	42^b	43
Basal Inhibition																	
17. GSR (1)	—	71	59	23	26	24	28	13	04	12	04	−05	−29	31	18	−08	12
18. GSR (2)		—	58	13	20	20	12	−04	17	11	20	−06	−34	21	00	02	21
21. GSR, generalization			—	16	14	37	38	12	12	20	01	−11	−24	16	06	11	32
46. Aniseikonic, time				—	31	15	−09	−24	14	22	06	−02	−13	13	13	01	14
47. Aniseikonic, tilt					—	04	01	−02	01	−02	−16	−07	−03	03	−03	−06	12
39. CFF, up-down						—	58	−20	04	16	−11	15	−07	01	08	11	05
40. CFF, median							—	68	19	14	−14	−03	−05	03	08	−19	04
Satiation																	
41. CFF, difference								—	19	02	−08	−18	00	03	03	−33	00
32. Spiral (1)									—	53	06	04	−11	−09	−13	−05	08
33. Spiral (2)										—	13	−12	−06	−12	04	−15	06
38. Necker, difference											—	−15	−05	08	−02	−19	−03
28. Kinesthetic AE (1)												—	04	−07	−06	−01	−02
31. Kinesthetic AE (2)													—	−29	−11	11	04
Reactive Inhibition																	
24. Rotor reminiscence (1)														—	44	−13	01
25. Rotor reminiscence (2)															—	−10	−02
42. Response alternation (random)																—	38
43. Response alternation (runs)																	—

[a] Signs have been changed so that a positive r is consistent with a general inhibition factor. An r of .21 is significant at the .05 level, one-tailed test.

[b] Variables of low retest reliability and/or construct validity.

general inhibition factor. However, Factor 3 is probably best considered a basal inhibition factor until more reliable measurement can indicate the limits of its generality.

Two other factors merit consideration. Factor 1 shows a consistent pattern of signs for all four reactive inhibition measures and also loads two satiation and two basal inhibition measures. As will be seen shortly, Factor 1 turns out to be primarily a sex difference factor and thus in terms of present theory, cannot be considered a function of a variety of inhibition. Factor 8, on the other hand, shows consistent loadings on the more reliable satiation measures. This too must be ruled out as a satiation factor on the basis of later evidence.

In summary, the evaluation of the experimental measures indicates: (*a*) the presence of a basal inhibition factor that may also include some aspects of satiation and reactive inhibition measures; (*b*) a failure to find any factor corresponding to Eysenck's concept of reactive cortical inhibition which includes satiation and reactive inhibition measures; and (*c*) a lack of convincing evidence for the existence of unitary traits fitting the definitions of satiation or reactive inhibition. Before leaving Table 10 to discuss each factor, the reader should note the distribution of factor loadings for the five random variables. Nine out of 40 factor loadings equal or exceed .25, the .05 significance level for r ($N = 62$). If factor loadings were distributed as r, only 2 out of 40 loadings would exceed .24. The high loadings of the random variables should serve to caution the reader against over-interpretation of small factor loadings, and should provide an empirical base for interpreting the significance of the factor loadings.

In the paragraphs to follow, each factor will be discussed briefly. Where necessary, salient variables will be presented in tabular form in such a way that the *description given corresponds to a positive factor loading*. The salient variables on factor 1 are as follows:

48. Female sex	74
11. Guilford M, Feminine interests	74
3. Cattell I, Premsia (feminine)	71
49. Young age	53
4. Cattell M, Alaxia (bohemian)	46
5. Cattell N, Naivete	46
46. Aniseikonic, short time delay	40
20. GSR, smaller UR to shock (2)	35
25. Rotor, more reminiscence (2)	30

None of the loadings on this *sex difference* factor are too surprising. The smaller reaction to shock on the GSR is most likely a function of

TABLE 10

A. FACTOR LOADINGS OF EXPERIMENTAL VARIABLES

Variables	Predicted Pattern A or B		Factor Loadings 1	2	3	4	5	6	7	8	H²
Basal cortical inhibition measures											
17. GSR (1)	−	+	0	0	67	0	+	−22	0	−23	58
18. GSR (2)	−	+	0	+	66	+	+	−	−20	0	56
21. GSR, generalization	−	+	0	+	72	−	0	−	−	0	58
46. Aniseikonic, time	−	+	−40	+	29	0	0	0	−21	+	33
47. Aniseikonic, tilt	+	−	+	0	−42	0	−	−	0	−	25
39. CFF, up-down	+	−	24	0	−48	+	+	0	0	0	34
40. CFF, median	+	−	0	0	−49	+	0	0	−50	+	53
Satiation measures											
41. CFF, difference	−	+	27	0	+	0	+	0	66	−20	59
32. Spiral (1)	−	+	−	36	0	0	22	27	0	−24	36
33. Spiral (2)	−	+	−21	39	+	0	0	+	0	−34	36
38. Necker, difference	+	−	0	−	+	−67	0	+	+	+	55
28. Kinesthetic AE (1)	+	−	+	0	0	+	31	−35	0	0	28
31. Kinesthetic AE (2)	+	−	0	0	28	0	25	−	−	−45	39
Reactive inhibition measures											
24. Rotor reminiscence (1)	+	−	23	+	−	0	0	57	0	+	45
25. Rotor reminiscence (2)	+	−	30	0	0	+	24	51	−30	0	52
42. Response alternation (random)	+	−	+	36	−	0	+	−23	38	−	41
43. Response alternation (runs)	+	−	+	26	−48	0	0	−26	0	0	39
Number of loadings over .25:											
Agreeing with Pattern A			2	2	1	0	1	2	3	1	
Agreeing with Pattern B			1	2	8	1	0	3	1	1	

B. FACTOR LOADINGS OF EXTRAVERSION, ANXIETY, AND RANDOM VARIABLES

Variables		1	2	3	4	5	6	7	8	H²
8. Rhathymia		−08	01	02	08	81	07	−06	03	68
12. Extraversion (Eysenck)		06	−06	11	−12	82	06	07	05	71
13. Neuroticism (Eysenck)		10	66	07	19	−28	07	03	−11	58
14. Anxiety (Cattell)		16	72	02	−16	−14	−15	02	−03	62
15. Extraversion (Cattell), based on the sum of factors A, E, F, & H		−02	−08	13	14	68	14	13	14	56
48. Sex, higher score male		−74	06	15	18	−08	−35	06	−08	74
54. Random Variable 1		13	−06	13	−09	−12	46	04	12	29
55. Random Variable 2		−01	39	−16	−23	05	11	−25	10	32
56. Random Variable 3		−25	−15	18	48	−02	22	06	−21	44
57. Random Variable 4		−09	−03	10	10	−18	03	−42	02	24
58. Random Variable 5		25	29	−14	00	31	−16	21	03	34

Note.—A zero in the table indicates a loading less than ±10; a + or − sign indicates a loading of 10 or greater but less than 20.

E's tendency to set easier shock levels for women than for men. The rotor reminiscence loading is probably a function of the poorer motor skill of women and the resulting greater build-up of interference effects or the greater effort required when skill is low. The shorter time delay with aniseikonic lenses is consistent with an earlier finding by the author and implies a need to control for such a sex difference in univariate studies.

Factor 2 has the following salient variables:

14. Cattell second-order, Anxiety	72
13. Eysenck N, Neuroticism	66
9. Guilford Co, Uncooperative	66
10. Guilford Ag, Disagreeable	64
5. Cattell N, Shrewdness	40
33. Spiral, longer afterimage (2)	39
55. Random Variable 2	39
32. Spiral, longer afterimage (1)	36
42. Response alternation, random series (1)	36

This factor is defined primarily by the Cattell and Eysenck Anxiety or Neuroticism scales. The loadings of experimental variables are largely inconsequential. It is suspected that the longer spiral afterimage scores represent mainly a response bias effect, i.e., the tendency for anxious subjects to agree, and therefore to continue to say that they see an effect even though they are not sure. Although the conditioning measure used is not the *most* appropriate for testing Spence's theory relating anxiety and conditioning (Farber, Spence, & Bechtoldt, 1957; Spence, 1958), it is interesting to note the failure of conditioning measures to load on this factor. The only GSR measure to approach a "significant" loading was UR to shock (.24).[7]

The salient variables loading on Factor 3 are as follows:

21. GSR, greater generalization	72
17. GSR, greater conditioning (1)	67
18. GSR, greater conditioning (2)	66
16. GSR, lower basal (1)	58
40. CFF, higher threshold, median of 1″ measures	49
39. CFF, higher threshold, up-down series	48
43. Response alternation, fewer alternations (run series)	48
47. Aniseikonic, less estimated tilt	42
37. Necker cube, fewer reversals, last 28 sec. (3 min. trials)	31
50. Father's education higher	30
46. Aniseikonic, longer time delay	29

Factor 3, as noted earlier, comes closest to any of the factors to representing a general cortical inhibition factor. From the predominance of conditioning, CFF, and aniseikonic lense loadings it is apparent that this factor reflects primarily a basal inhibition effect, rather than a reactive inhibition or satiation effect. Of greatest importance is the failure to find *any* personality measures associated with this factor.

Factor 4 does not require much elaboration. It loads in the 70s on fast alternating perspective as measured by the 1-min. Necker cube trials and by the first 28 sec. of the 3-min. Necker cube trials. In addition, it loads greater UR to shock on the GSR (.51 and .43). Finally, *S*s showing the greater neural reactivity suggested by the above loadings are found to get more hours of sleep per night (.44).

Factor 5 requires closer scrutiny. The salient variables loading on Factor 5 are as follows:

12. Eysenck E, Extraversion	82
8. Guilford R, Rhathymia	81
15. Cattell Second-order Extraversion	68
7. Guilford T, Thinking Extraversion	61
5. Cattell G, Low Super Ego Strength	39
26. Kinesthetic figural aftereffect, higher baseline (1)	34
20. GSR, greater average UR (2)	34
28. Kinesthetic figural aftereffect, greater percentage decrement (1)	31
58. Random variable 5	31
29. Kinesthetic figural aftereffect, higher baseline (2)	25
31. Kinesthetic figural aftereffect, greater percentage decrement (2)	25
25. Rotor reminiscence (2)	24

Factor 5 is clearly an extraversion factor, loading all of the criterion personality measures. The only critical inhibition variable to load over .30 on this factor is the kinesthetic aftereffect decrement score (1) which was found to have zero retest reliability and to correlate opposite to hypothesis with the other inhibition measures (Table 9). There are several other reasons why this decrement score loading cannot be considered supportive evidence for Eysenck's hypothesis. First, the correlations of the decrement score with extraversion measures are on the order of .12, quite insignificant. Second, one of the random variables loads as highly. Third, the baseline kinesthetic aftereffect measure loads just as high as the decrement score, suggesting the possibility that any found association between extraversion and the decrement score is due solely to initial baseline difference. The present correlations support this latter hypothesis, though a further study is required to clarify it. The reason for reporting the small retest loading of the rotor reminiscence measure will become apparent later.

Factor 6 appears to be a motor skill factor associated with sex differences. It loads the measures of pursuit rotor skill in the 50's and both reminiscence measures negatively, indicating that less reminiscence occurs with a higher level of skill. Future attempts to measure reactive inhibition as an individual difference are going to have to find some way to eliminate this confounding of reactive inhibition with motor skill.

[7] A more appropriate test of Spence's theory, in contrast to Eysenck's, would use the magnitude of conditioned response corrected for basal artifact. Such tests also failed to reveal any significant relationships to anxiety measures. However, in a GSR conditioning study carried out subsequent to the writing of this paper, Becker and Matteson (in press) used extreme groups on anxiety and extraversion dimensions and found confirmation for Spence's hypothesis, but failed again to find confirmation for Eysenck's hypothesis.

Factor 7 is interesting because of its coloading of personality and objective test measures. Cattell's Q_1, Radicalism, loads .55 and Guilford's T, Introspective, loads .31. The main loading of objective tests are on higher CFF threshold (one second exposures) and on reaction time measures. On the objective test side this factor is quite similar to Cattell's U.I. (T) 22 which he calls Corticalertia or Nervous, Alert Reactivity.

Finally Factor 8 appears to be the remaining variance of the general tendency to overestimate the size of the standard test object under both baseline and postsatiation test conditions in the kinesthetic aftereffect experiment. The only personality test measure found to load on this factor is Cattell's B (IQ) which loaded .35. It is very difficult to say just what B is measuring in the particular context of personality test items in which it is given. The present writer finds zero retest reliability of parallel form measures of B with college students.

Correlational Analysis

Although factor analysis can frequently reveal relationships not apparent in a correlation matrix, factor analysis is handicapped by the difficulty one has in stating the significance of a factor loading. One approach to this problem was the inclusion of random variables in the factor analysis. Another approach is the direct evaluation of the probable significance of the correlations between the 17 inhibition and three extraversion measures.

Using a one-tailed test at the .05 level, there are only two significant *r*'s between the 17 inhibition measures and the three extraversion measures. Since both of these significant *r*s involve the same inhibition measure (rotor reminiscence, retest) and are with Eysenck's E and R, even these two significant *r*s are not independent. On a statistical basis this is a finding that would occur with a high probability on a chance basis. Thus, on statistical grounds, one is forced to the same conclusion apparent from the factor analysis, namely, that *there is no evidence for a relationship between any kind of inhibition measure* (basal inhibition, satiation, reactive inhibition) *and extraversion.*

DISCUSSION

Methodological Problems

The evidence of the present study indicates that it is particularly difficult to get reliable measures of individual differences in *satiation* and *reactive inhibition* effects. If the within sessions reliabilities for these measures were high and the between sessions reliabilities were low it might be possible to argue that one is dealing with very unstable variables which show considerable within individual fluctuation. However, since the present within-session reliabilities were also low (.50 or less), it is not yet possible to decide whether the problem is one of learning how to get more reliable measures or whether there are no consistent individual differences in satiation and reactive inhibition effects.

The kinesthetic figural aftereffect device has methodological difficulties connected with it which neither Klein and Krech (1952) nor Eysenck (1955) have attempted to take into account. Unlike the others, the present *E* took retest measures on his *S*s. It was found that mean baseline dropped from 79 to 76 (objective equality equals 73). Using a test for the significance of difference between correlated means, this difference is significant beyond the .01 level. This finding suggests a need to find out just how much decrement would occur from simply taking repeated judgments of the comparison stimulus with no satiation involved at all. It remains a distinct possibility that the relationship of satiation to extraversion found by Eysenck with this device is solely a function of greater initial overestimation by extroverts and subsequent approaching of more objective judgment with practice.[8] Another methodological problem with the kinesthetic figural aftereffect device is the method used in scoring a satiation effect. The percentage decrement score is a very complex function involving both the complexities inherent in a difference score and those inherent in a ratio score. A more directly meaningful approach to establishing relationships to such satiation measures would be to use the partial correlation method discussed earlier. Other statistical problems in

---

[8] In a subsequent study, Northrup and Becker (in preparation) failed to find a relationship between extraversion and baseline judgments or satiation decrements. Also see Footnote 3.

Eysenck's original report using this device (1955) have been adequately covered by Rechtschaffen (1958) and need not be repeated here except to note that proper statistical analysis would not have led to the conclusion that the results reported by Eysenck were statistically significant. Finally, it should be noted that Crutchfield, Woodworth, and Albrecht (1958) found that the amount of satiation effect produced in the kinesthetic figural aftereffect apparatus correlated .52 with the tendency for satiation effects to dissipate over time. This is contrary to Eysenck's hypotheses that extraverts should build up satiation effects more quickly and dissipate them more slowly.

Reactive Inhibition and Extraversion

It will be remembered that the only "significant" correlations between inhibition and extraversion measures were between the retest reminiscence score and both E and R. The *r*s were .21 and .22 respectively (and .19 with Cattell's Extraversion). Also, this retest reminiscence score loaded .24 on the Extraversion Factor. In view of Eysenck's previous replications, perhaps one should not be quite so hasty in attributing the finding to chance. At one point the writer suspected that the relationship might be attributable to sex differences or to motor skill differences since both of these variables correlated highly with the reminiscence score. Partialing out each of these variables failed to have any effect on the found relationships to the extraversion measures. It is also significant that the found relationship was to the second reminiscence measure, where more *S*s would be approaching their learning asymptotes (after an additional 16 min. of practice) and thus the reminiscence score would more likely be reflecting reactive inhibition rather than other factors associated with the learning process. In reviewing this result, consideration should also be given to Rechtschaffen's (1958) finding of an *r* of .18 between R and an inhibition measure based on the difference in performance on a motor task under massed versus distributed conditions. The above mentioned *r* would be significant at the .05 level, using a one-tailed test, if it were the only relationship being considered. However, Rechtschaffen was in the position of comparing a number of relationships and on statistical grounds the findings could be attributed to chance.

The accumulation of a number of such correlations of borderline significance suggests that Eysenck's results may be replicable. Opposing the acceptance of this conclusion are several other facts. First, the response alternation measures (based on runs) which behaved in accordance with reactive inhibition theory failed to relate to reminiscence or extraversion. Rechtschaffen's (1958) reminiscence measure failed to relate to R. In studies preliminary to this one, the writer failed to find significant differences between extreme extroverts and extreme introverts on pursuit rotor measures under massed and distributed conditions. Also, Miller (1957) failed to find any relationship between R and reminiscence or the difference between massed and distributed conditions on a serial-rote learning task. In the face of these facts, it must be concluded that if a relationship between reactive inhibition and extraversion exists, it is probably of such a low magnitude to be hardly worthy of further study.

Other Related Studies

In weighing the evidence for or against the hypotheses under consideration, several other results should also be considered. In the work preliminary to this study which included rotor reminiscence, visual satiation, aniseikonic lenses, and the Necker cube, no significant relationships to R were found. Farber, Spence, and Bechtoldt (1957) failed to find any significant relationships between eyelid conditioning and R. Rechtschaffen (1958) failed to find any relationship between a Kohler-type visual satiation measure and R. Using the kinesthetic figural aftereffect satiation measure, Crutchfield, Woodworth, and Albrecht (1958) report that among officers greater satiation effects are associated with staff ratings of passivity, suggestibility, lack of dominance, and lack of leadership skills. While the magnitude of these relationships are no larger than those found by Eysenck (.20 to .30), it is of note that they are in *exactly the opposite direction* from those which Eysenck's theory would predict. Finally, mention should be made again of the study by Meier

(1956) in which he found a significant negative correlation between the amount of reminiscence in inverted alphabet printing and the amount of kinesthetic aftereffect. (The *r*s in the present study between kinesthetic aftereffect and rotor reminiscence measures were all negative, ranging from −.06 to −.29).

SUMMARY AND CONCLUSIONS

This study set out to examine Eysenck's hypothesis that cortical inhibition is one of the bases for individual differences in extraversion-introversion. From a review of the literature it became apparent that three distinct concepts were involved in this hypothesis in one way or another. First, Eysenck subsumed both satiation effects and reactive inhibition effects under the cortical inhibition concept. Second, it became apparent that Franks' use of cortical inhibition was quite distinct from Eysenck's. Franks used the concept of cortical inhibition to refer to a basal trait-like difference among individuals which was independent of the number of stimulations of a given neural structure. Thus, in effect, it was found that three different hypothetical variables, all going under the name of cortical inhibition, were assumed to be related to individual differences in extraversion-introversion. The purpose of this study was to find out which, if any, of these three constructs (satiation, reactive inhibition, basal cortical inhibition) could be subsumed under a more general construct, and to find out which, if any, of these constructs were related to individual differences in extraversion-introversion.

Sixty-two college students of both sexes randomly selected from an elementary psychology pool of 1000 students were used as *S*s. Consistent with Eysenck's earlier work, criterion measures of extraversion-introversion were derived from the Guilford and Cattell personality inventories. Each *S* participated in five hours of laboratory testing in addition to the personality testing. A variety of experimental tests were designed or selected which would produce scores interpretable as measures of basal cortical inhibition, satiation, or reactive inhibition. While it was not always possible experimentally to separate basal and reactive effects, two of the eight experimental tests used clearly measured reactive inhibition effects (pursuit rotor reminiscence and response alternation), three fitted the definition of a satiation effect (kinesthetic aftereffect, Archimedes spiral, and Necker cube difference score), while three approximated basal type measures (GSR conditioning, aniseikonic lenses, and CFF). Some exceptions to these general classifications are noted in the text. Including retest measures taken for many of the variables a week or two later, 17 experimental measures were considered critical for the evaluation of the hypotheses.

The findings indicated that the concept of basal cortical inhibition as used by Franks appeared as a unitary factor, but was not found to be related to extraversion-introversion. No empirical evidence was found to support Eysenck's assumption that satiation and reactive inhibition form a unitary trait. Satiation and reactive inhibition measures were found to have some common variance with the basal inhibition measures, but they did not covary with each other. There was no evidence to support Eysenck's hypotheses that satiation measures covary with extraversion measures. There is very minimal evidence that some reactive inhibition measures covary with extraversion. However, there is equally compelling evidence to the contrary. If a relationship between reactive inhibition and extraversion exists, it is probably of such a small magnitude as to be practically and theoretically trivial.

REFERENCES

AMMONS, R. B. Acquisition of motor skills: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.

BECKER, W. C. Perceptual rigidity as measured by aniseikonic lenses. *J. abnorm. soc. Psychol.*, 1954, **49**, 419–422.

BECKER, W. C. Cortical inhibition and extraversion-introversion Unpublished final rep., *USPHS Project No.* M-1713, Univer. Illinois, 1959.

BECKER, W. C., & MATTESON, H. H. GSR conditioning anxiety, and extraversion. *J. abnorm. soc. Psychol.*, in press.

CATTELL, R. B. *Personality and motivation structure and measurement.* New York: World Book, 1957.

CRUTCHFIELD, R. S., WOODWORTH, D. G., & ALBRECHT, RUTH E. Perceptual performance and the effective person. *USAF tech. Note*, 1958, No. 58–60.

EYSENCK, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.

EYSENCK, H. J. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, **53**, 328–333.

FARBER, I. E., SPENCE, K. W., & BECHTOLDT, H. P. Emotionality, introversion-extraversion, and conditioning. Paper read at Midwestern Psychological Association, Chicago, May, 1957.

FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **2**, 143–149.

FRANKS, C. M. Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 1957, **48**, 119–126.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

KIMBLE, G. A., & HORENSTEIN, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.

KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, **21**, 118–148.

KÖHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

LANDIS, C. Something about flicker-fusion. *Scient. Mon.*, 1951, **73**, 308–314.

MCNEMAR, OLGA W. The ordering of individuals in critical flicker frequency under different measurement conditions. *J. Psychol.*, 1951, **32**, 3–24.

MEIER, M. J. Interrelationships among personality variables, kinesthetic figural aftereffects, and reminiscence in motor learning. Unpublished doctoral dissertation. Univer. Wisconsin, 1956.

MILLER, PAULA E. Reactive inhibition in extraversion and introversion. Unpublished bachelor's thesis, Univer. Illinois, 1957.

PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffects. *J. consult. Psychol.*, 1955, **19**, 299–302.

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283–291.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

(Received June 19, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 67–72

CHANGES IN SELF-IDEAL DISCREPANCY IN REPRESSORS AND SENSITIZERS[1]

JOHN ALTROCCHI, OSCAR A. PARSONS,[2] AND HILDA DICKOFF
Duke University

RESEARCH by Rogers and his colleagues has suggested that the concepts of self and ideal-self become less discrepant as a function of brief psychotherapy and that such changes reflect improvement in adjustment (Butler & Haigh, 1954). Other investigators, such as Block and Thomas (1955) and Hatfield (1958), have found evidence of a curvilinear relationship between psychiatric adjustment or personality integration and the discrepancy between self and ideal-self. The latter investigators also demonstrated that persons with very low self-ideal discrepancies tend to use repressive and denying defenses. The purpose of this study is to attempt further specification of the variables related to self-ideal discrepancies by investigating whether people with different defenses or modes of adaptation differ in self-ideal discrepancy and change in different ways subsequent to psychotherapeutic interpersonal interaction.

The mode-of-adaptation dimension which is suggested is repression-sensitization, the terms having been applied by Gordon (1957) to a dimension suggested by Eriksen and his colleagues (Lazarus, Eriksen, & Fonda, 1951) from research in perceptual defense. Repressors are defined as those who tend to use avoidance, denial, and repression of potential threat and conflict as a primary mode of adaptation. Sensitizers are defined as those who are alerted to and perhaps overinterpret potential threat and conflict and who use intellectual and obsessive defenses as a primary mode of adaptation. It has been shown that people are consistent in their use of either repression or sensitization across task situations (Lazarus et al., 1951; Eriksen, 1952) and that repressors, compared to sensitizers, recall more completed than incompleted tasks under threat (Eriksen, 1954), avoid the expression of unacceptable needs on projective tests and have higher perceptual thresholds for material related to such needs (Eriksen, 1951), recall fewer disturbing words (Carlson, 1954), forget in response to implied failure in a learning task (Truax, 1957), and predict that partners' responses will be more similar to their own responses on a personal inventory (Gordon, 1957).

Since repressors tend to remain unaware of threat, conflict, and negative attitudes and to focus on the positive qualities of themselves and others (Leary, 1957), they can be expected to manifest a small discrepancy between self and ideal-self. Since sensitizers are attuned to and tend to ruminate about threat, conflict, and the negative qualities of themselves and others (Gordon, 1959), they can be expected to manifest a high self-ideal discrepancy. Therefore:

Hypothesis 1—Repressors manifest smaller self-ideal discrepancies than sensitizers.

Brief psychotherapy or a similar kind of experience aimed toward personality growth and increased insight into one's behavior with others should cause the repressors to begin to recognize negative aspects of themselves; their self-ideal discrepancies should therefore increase. Such an experience should cause the sensitizers to begin to blame themselves less and to focus less on their negative and more on the positive qualities, and their self-ideal discrepancies should decrease. Therefore:

Hypothesis 2—As a result of training focusing on psychotherapeutic interpersonal interaction, the self-ideal discrepancies of repressors increase and the discrepancies of sensitizers decrease.

---

[1] An earlier version of this paper was presented at the meetings of the APA in August, 1958. This study was carried out within the context of and was partially supported by Grant M-1599 from the National Institute of Mental Health. Data analysis was also partially supported by grants from the Duke University Research Council. The authors wish to express their gratitude to Gerald Levin and Tallulah Brown for some of their work in the earlier phases of this study and to Robert Dovenmuehle, Faye Spring, and Betty Sue Johnson, co-investigators in the larger project, for their cooperation.

[2] Now at the University of Oklahoma Medical Center.

Subjects in general, however, disregarding their place on the repression-sensitization dimension, should show the effect found by Rogers and his colleagues (Butler & Haigh, 1954), a decrease in self-ideal discrepancy. Therefore:

Hypothesis 3—As a result of training focusing on psychotherapeutic interpersonal interaction, subjects' self-ideal discrepancies decrease.

METHOD

Eighty-eight senior student nurses at the Duke University School of Nursing in 1956–57 composed the experimental (E) group, and 64 senior student nurses in 1957–58 made up a cross-validation (CV) group. *Ss* were tested at the beginning and end of a 16-week period of training in psychiatric nursing. This intensive training was interpersonally and psychodynamically oriented and included individual counseling aimed at helping the student become more aware of the dynamics underlying her own behavior in her relationships with patients. The assumption was made that such training tends to facilitate emotional growth and enhance self-insight in somewhat the same manner as brief psychotherapy. This assumption is supported by the subjective reports of these student nurses and their instructors, by anecdotal evidence in clinical psychology and related fields, and by some indirect research evidence. For instance, Bernstein, Brophy, McCarthy, and Roepe (1954) found that nurses undergoing a similar experience shifted in their attitudes toward interpersonal relationships.

An attempt was made to develop a more detailed measure of the independent variable, the repression-sensitization dimension. For repressors, the most recent studies used the Denial subscale (*Dn*) rather than the whole *Hy* scale of the MMPI (Gordon, 1959; Hatfield, 1958). Because of the shortness of the *Dn* scale and to be consistent with Gordon (1957), the Lie (*L*) and Defensiveness (*K*) scales were also used to select repressors.[3] The intercorrelations of *L*, *K*, and *Dn* range between .32 and .78 (Little & Fisher, 1958; Hatfield, 1958). For sensitizers, the *Pt* scale has traditionally been used. Since *K* and *Pt* correlate negatively, however (Hatfield, 1958), *Pt* was used without the addition of *K*. Again, to lessen the likelihood of measurement error, the Depression (*D*) and Anxiety (*A*) (Welsh, 1956) scales also were used. Thus, the *T* scores of *D*, *Pt* (without *K*), and *A* were subtracted from the *T* scores on *L*, *K*, and *Dn*.[4] In each group, the 15 *Ss* with the highest positive scores were called repressors and the 15 with the highest negative scores were called sensitizers. For comparison purposes, a middle group was included.

TABLE 1

MEAN SELF IDEAL DISCREPANCIES AND CHANGES IN THE E GROUP

	Pre-training	Post-training	Change	*t* for Related Means	*p*
Repressors ($N = 15$)	24.7	27.5	−2.8	0.7	*NS*
Midgroup ($N = 15$)	36.1	32.1	4.0	1.7	*NS*
Sensitizers ($N = 15$)	54.0	54.3	−0.3	0.2	*NS*
Total Group ($N = 88$)	40.6	37.3	3.3	2.0	$<.05$

Self-ideal discrepancy was assessed by the use of the Interpersonal Check List (ICL) (Leary, 1957), which consists of 128 phrases to be checked true or false, 16 phrases falling into each of 8 octants, each reflecting a different kind of interpersonal behavior. Discrepancies were calculated by adding across octants the absolute discrepancies between self and ideal-self.

RESULTS

The distributions of repression-sensitization scores in the E and CV groups were investigated. Repressors in the E group had scores ranging from +49 to +95 with a mean of +67. Repressors in the CV group ranged from +40 to +80 with a mean of +54 (difference significant, $p < .01$, *U* test). For sensitizers the scores were as follows: E group, −11 to −78 with a mean of −37; CV group, −18 to −116, with a mean of −51 (difference not significant). On this dimension, then, the two groups were not identical.

Tables 1, 2, and 3 show that Hypothesis 1 was clearly supported.[5] Repressors had much smaller self-ideal discrepancies than sensitizers in the E group and in the CV group. Furthermore, this difference was still apparent after training. In every case, the middle group fell between the repressors and sensitizers.

Hypothesis 2, that repressors and sensitizers change differentially in self-ideal discrepancy after training, was not supported in either group (See Tables 1 and 2). There did not, in fact, seem to be any consistency to the changes of repressors or sensitizers.

---

[3] Welsh's Repression (*R*) scale was not used because the construction and correlates of this scale (Welsh, 1956) suggest a variable different from the repression referred to here.

[4] The investigators wish to express their gratitude to G. S. Welsh, W. G. Dahlstrom, and H. G. Gough for their advice concerning the measurement of repression-sensitization.

[5] The lower scores in Table 2 result from shifting from Leary's original weighting system to his later simplified scoring, where equal credit is given to every item. Since no published comparison of the two weighting systems was found, the investigators examined the relationships in a small sample and found correlations of .90 to .95.

Hypothesis 3 received partial support. The self-ideal discrepancy of the E group decreased significantly, but a similar trend did not reach an acceptable level of significance in the CV group (See Tables 1 and 2).

TABLE 2

MEAN SELF-IDEAL DISCREPANCIES AND CHANGES IN THE CV GROUP[5]

	Pre-training	Post-training	Change	t for Related Means	p
Repressors ($N = 15$)	16.0	13.9	2.1	1.0	*NS*
Midgroup ($N = 15$)	19.9	18.6	1.3	0.7	*NS*
Sensitizers ($N = 15$)	22.2	21.7	0.5	0.3	*NS*
Total Group ($N = 64$)	18.4	17.3	1.1	1.4	*NS*

TABLE 3

TESTS FOR DIFFERENCES IN SELF-IDEAL DISCREPANCIES PRE- AND POSTTRAINING

	E Group				CV Group			
	Pre-training		Post-training		Pre-training		Post-training	
	t	p	t	p	t	p	t	p
Repressors vs. Sensitizers	5.16	<.001	4.12	<.001	2.59	<.05	2.63	<.02
Repressors vs. Midgroup	2.40	<.05	.94	*NS*	1.72	*NS*	2.03	<.10
Midgroup vs. Sensitizers	2.81	<.02	3.17	<.01	.83	*NS*	.84	*NS*

In order to understand better how repressors and sensitizers differ in self-ideal discrepancy and to discover whether any reliable changes occurred during this training, a number of further analyses were undertaken. To delineate the components of the highly significant differences in self-ideal discrepancy, octant scores were analyzed. No significant differences were found between repressors and sensitizers on any octant for ideal-self. Table 4 shows that the difference between repressors and sensitizers in self-ideal discrepancy was primarily accounted for by some rather specific differences in self-concept. Sensitizers consistently described themselves as more rebellious-distrustful and, with somewhat less consistency, as more aggressive-sadistic and more self-effacing-masochistic.

In order to discover whether any reliable changes occurred in specific dimensions of self- or ideal-self-concepts and whether any such changes occurred differentially in repressors and sensitizers, analyses of variance of octant scores were attempted. The results suggested that the octant scores were not stable enough for these complex analyses. Therefore, analyses were made of scores that are less specific but presumably more stable—the Love (Lov) and Dominance (Dom) scores, derived by combining octants (Leary, 1957). (See Tables 5 and 6.) The only result which was significant in both groups was that repressors described themselves as more dominant before and after training. This result merely highlights what was revealed earlier (See Table 4): repressors and sensitizers differed in

TABLE 4

PROBABILITY VALUES FROM t TESTS FOR DIFFERENCES BETWEEN REPRESSORS AND SENSITIZERS IN SELF-CONCEPT BY OCTANTS OF THE INTERPERSONAL CHECK LIST, PRE- AND POSTTRAINING[a]

Octant	Description	E Group		CV Group	
		Pretraining	Posttraining	Pretraining	Posttraining
AP	Managerial–Autocratic				
BC	Competitive–Narcissistic				
DE	Aggressive–Sadistic	<.01	<.05		<.10
FG	Rebellious–Distrustful	<.001	<.001	.01	<.01
HI	Self-Effacing–Masochistic	<.01	<.10		<.01
JK	Docile–Dependent				
LM	Cooperative–Overconventional				
NO	Responsible–Hypernormal				

[a] In all significant comparisons, sensitizers had higher scores.

discrepancies of sensitizers decrease, was not confirmed. Hypothesis 3, that subjects' self-ideal discrepancies decrease, was partially supported. Some changes in self or ideal-self, especially on the Love dimension, were found and these changes were consistent with the goals of the training, but the effects of the training from one year to the next were not entirely consistent. Studying the components of the self-concept was found to be more fruitful than studying the gross self-ideal discrepancy.

REFERENCES

BERNSTEIN, L., BROPHY, MARIE L., MCCARTHY, MARY JANE, & ROEPE, RUBY L. Teaching nurse-patient relationships: An experimental study. *Nurs. Res.*, 1954, **3**, 80–84.

BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254–259.

BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954. Pp. 55–75.

CARLSON, V. R. Individual differences in the recall of word-association test words. *J. Pers.*, 1954, **23**, 77–87.

ENGEL, MARY. The stability of the self-concept in adolescence. *J. abnorm. soc. Psychol.*, 1959, **58**, 211–215.

ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–564.

ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230–235.

ERIKSEN, C. W. Psychological defenses and ego strength in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.

EYSENCK, H. J. *Dimensions of personality.* London: Routledge, Paul, 1947.

GORDON, J. R. Interpersonal predictions of repressors and sensitizers. *J. Pers.*, 1957, **25**, 686–698.

GORDON, J. R. The stability of the assumed similarity response set in repressors and sensitizers. *J. Pers.*, 1959, **27**, 362–373.

HATFIELD, J. A study of self-concept configurations in relation to forms of ego functioning. Unpublished doctoral dissertation, Univer. California, 1958.

LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–482.

LEARY, T. *Interpersonal diagnosis of personality.* New York: Ronald, 1957.

LITTLE, K. B., & FISHER, J. Two new experimental scales of the MMPI. *J. consult. Psychol.*, 1958, **22**, 305–306.

SMITH, G. M. Six measures of self-concept discrepancy and instability: Their interrelations, reliability, and relations to other personality measures. *J. consult. Psychol.*, 1958, **22**, 101–112.

TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. consult. Psychol.*, 1955, **19**, 205–209.

TRUAX, C. B. The repression response to implied failure as a function of the hysteria-psychasthenia index. *J. abnorm. soc. Psychol.*, 1957, **55**, 188–193.

WELSH, G. S. Factor dimensions *A* and *R.* In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956. Pp. 264–281.

(Received June 30, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1 73-81

EFFECTS OF GROUP SUPPORT ON THE EVALUATION OF AN ANTAGONIST[1]

LLOYD H. STRICKLAND[2]
University of North Carolina

EDWARD E. JONES
Duke University

AND WILLIAM P. SMITH
University of North Carolina

Our impressions of others are determined not alone by information which these others directly provide; the verbal and gestural responses of a "stimulus person" (*SP*) are typically assessed by a perceiver in relation to the situation in which they occur (Jones & Thibaut, 1958). In many cases, the perceiver himself is a component of this situation. He must therefore consider his own behavior as a condition affecting the other's responses and the meaning he assigns to them. If A, for example, is aware of his own role in the instigation of hostile behavior in B, he is less likely to think of B as hostile or unpleasant, than he would if he were unaware of the provocation he himself provided.

It seems to follow that, in the typical social interaction, evaluations of others on the basis of their responses are contingent on the perceiver's evaluation of his own behavior. If the perceiver behaves in a "good" fashion and this elicits a "bad" response (e.g., a negative evaluation, a frown, a critical comment) from the *SP*, the latter will be judged as personally bad by the perceiver. If the perceiver is unconvinced that his own behavior is good, worthy, or appropriate, he will be less likely to form a negative evaluation of *SP*. This line of reasoning suggests an experimental situation in which one person (the *SP*) responds with hostility to the behavior of another (the perceiver) which is or is not highly valued by the perceiver himself.

A potential determinant of this valuation is the extent to which the perceiver's initiating response has normative support in some relevant group. A hostile reaction to behavior that has such normative support would seem to invite retaliation in the form of a negative impression. Hostile reaction to the same initiating response unsupported by social sanction would result in a less derogatory impression of the hostile respondent.

In simplified form, this is the reasoning behind the main prediction of the experiment described below. Consider an experimental procedure involving three primary steps or stages. In Step I, one group member (*S*) chooses influence messages to transmit to the target person or *SP* (not a group member); in Step II *S* learns that others in the group do (Group Support) or do not (No Support) endorse the particular messages selected; in Step III, *S* hears the *SP* severely criticize his messages and derogate their source. In such a situation, the main experimental hypothesis is that *perceivers in the group support condition will be more negative in their final impressions of the SP than will perceivers in the no support condition.*

Although it is theoretically quite possible that *S*s could respond to such a situation by rejecting their group in the No Support condition or succumbing to various forms of withdrawal and perceptual distortion, the experiment was designed to minimize such alternative reactions and thus to promote differential perceptions of the *SP*. Several post-experimental measures failed to reveal any tendency for *S*s in either group to adopt these different ways of managing the problem created by the *SP*'s hostility. Therefore, these alternative reactions will not be considered further.

---

[1] This study was initiated in the program of the Organization Research Group of the Institute for Research in Social Science at the University of North Carolina. The ORG is supported by the Office of Naval Research Nonr-855(04). Analysis and writing time was made available to Jones on National Science Foundation contract G8857 administered by Duke University.

[2] Now at Dartmouth College.

METHOD

Procedure

Ninety male *S*s, recruited from introductory psychology courses, were observed in groups of three. Each group was told that the experiment was part of a broader research project concerned with opinion change and reactions to propaganda. The experimental rationale emphasized that any propagandist must be aware of the "characteristics of his audience," which could be inferred by the audience's response to an influence attempt. *S*s were told that one of their group would eventually present a number of persuasive arguments about "big time athletics" to a fourth student (who would be in an adjacent, sound-proof room) while the two remaining *S*s would be "innocent bystanders." Their job was two-fold: (*a*) evaluate the persuasive arguments to be used; (*b*) observe the reactions of the fourth student and make some inferences about "just what kind of a person you think he is." It was explained to the group that the issue of "big time athletics" on the college campus had been chosen as the influence topic because of the differences in opinion on this subject, and that the *S*s had been selected to participate in the experiment because they had all had about the same scores on an Attitude-Toward-Big-Time-Athletics questionnaire, which they had filled out in class early in the semester. An attempt actually was made to compose groups of persons whose scores were relatively homogeneous and extreme in the *pro* or *con* direction, although it was not always possible to compose the groups of extremes. The fourth student was presented as one whose questionnaire responses reflected that he was "more or less undecided on the issue".

E stated that he would first interview the fourth student (henceforth referred to as the *SP*) in the next room. This interview would be carried through a microphone to a loudspeaker in the laboratory, so that the three *S*s might obtain an initial impression of him which would aid in their selection of good arguments supporting their own position. *E* left the room and turned on a previously recorded taped interview.

In this initial interview, an attempt was made to create a rather neutral but mildly likeable *SP*. He was presented as a sophomore, undecided on his major, who enjoyed intramural sports but was concentrating on his studies. He stated that he didn't care one way or the other about "big time" sports—he enjoyed going to ball games but didn't think sports "should push everything out of the picture."

After hearing this brief interview, the *S*s filled out a 10-item rating scale designed to reveal their "first impressions" of *SP*. They then were asked to bring up in a brief discussion those points (favoring their own position) which might be successful in influencing the interviewee. This was done in order to involve the *S*s more strongly in the issue and to increase the saliency of group membership by stressing attitudinal similarity and the cooperative nature of the task of finding arguments. Each *S* was then seated in his own private cubicle and assigned a letter (supposedly, *A*, *B* or *C*) for identification. The ostensible purpose of this procedure was to preserve the anonymity of the responsible *S*, whose arguments were to be transmitted to *SP*, vis-à-vis the innocent bystanders, who would be evaluating *S*'s arguments. Actually, all three *S*s received *B* identification cards.

Each *S* then selected from a list of 13 typical, one-paragraph arguments, the five which he thought most likely to influence *SP*, after having rated each of the 13 on a six-point scale from Poor to Excellent. He made these choices by placing a checkmark in a column headed B, in the left-hand margin of the pages, beside those preferred arguments. *E* then collected these sheets, supposedly to record *S*'s choices and to check in the adjoining A and C columns the choices of *B*'s partners. They were then returned to *S*.

In this way, the major independent variable was manipulated: in the Group Support condition, each *S* (*B*) perceived that *A* and *C* had agreed with his choices in four of five cases. In the No Support condition, *S* found his partners in close agreement with each other but in slight agreement with him.

Following the return of these argument selection sheets to the *S*s, *B* was formally selected (presumably by lot) as the letter designation of the Communicator, leaving the nonexistent *A* and *C* to function as Innocent Bystanders. All *S*s were instructed to select, from a pile of hand-written copies of the arguments, the five that *B* chose. *B* was told to rank order them in terms of their probable increasing effectiveness—*A* and *C* were to write brief evaluative comments about them. The rank-ordered arguments were collected from each *S*, and *E* announced that he would bring them to *SP*, let him read them, and then interview him about his reactions to them.

In the next interview, also prerecorded on tape, *SP* appeared initially vague and evasive. As he reviewed each argument (in *S*'s designated order), he became increasingly derogatory, not debating the point made by the argument, but disparaging the attempt to change anyone's mind with such poor material. After completing his review of the arguments, *SP* made a few additional unkind remarks about *S* and then asked *E* if he could leave, preferring not to meet his fellow *S*s. After granting this permission, *E* returned to the adjoining laboratory and required the *S*s to complete the following:

1. A brief, free-response personality sketch of *SP*, and a statement of arguments *S* might use were he to attempt to influence *SP* again.
2. A second, private rating of *SP* on the same scale form used in the initial impression rating.
3. A re-rating of all 13 arguments on the original six-point poor–excellent scale.
4. A "cohesiveness" measure, in which *S* designated (*a*) how much he would like to work with his fellow group members again, and (*b*) how much he felt they would prefer him as Communicator again.
5. A "public" rating of *SP*, on a parallel form of the scale used in the first and second ratings, purportedly to serve as the basis for a group discussion about *SP* and final composite group rating of him. It was expected that the predicted differential change in the private ratings would be augmented in making the public ratings.

6. A recognition-recall measure, in which *S* attempted to designate those five arguments initially preferred by himself and his two partners.

After these procedures were completed *E* informed the *S*s about the method and purposes of the experiment and determined the extent to which their suspicions had been aroused.

Rating Scale

As indicated above, each *S* rated the *SP* on three different occasions, before the derogation, after the derogation (private), and once again (public). Since the final two ratings were made consecutively, it was necessary to use two equivalent forms of the scale. In order to insure that the actual positive or negative results obtained were not solely a function of the particular items comprising the scale, approximately half of the *S*s were given Form I to use on the first and second rating and Form II to use on the third (i.e., public) rating. The order of form administration was reversed for the remaining *S*s. Each form of the scale consisted of 5 trait pairs (or 10 items) which measured the perceived characteristics: likeability, adjustment, warmth, intelligence, and conceit. The ratings were made on a six-point scale reflecting the degree of agreement with simple declarative statements about the person being rated. The scale was constructed in such a way that a positive statement like: "He seems to be a very warm and affectionate person" was in all cases matched by its negative counterpart in a different item such as: "He strikes me as a hostile and unsympathetic person." Many of the items on both forms of the scale were taken from a very similar scale described by Jones, Hester, Farina and Davis (1959). Data on reliability are included in the earlier study, but in view of the way in which the scales were used in the present study, separate reliability analyses were run on each scale. The split-half correlations of the change scores on Forms I and II were .57 and .61 respectively (corrected by the Spearman-Brown prophecy formula). Obviously, the scales each measure other variables in addition to overall favorability or "halo," but the reliability coefficients seem adequate to provide a meaningful test of changes in general evaluation of the *SP*.

Design and Distribution of Subjects

Although primary interest lies in the effects of the manipulation of group support, the two additional control variables make the design somewhat more complex. As Table 1 indicates, the design involved three dichotomous variables resulting in eight separate cells of *S*s. Table 1 also indicates the number of *S*s actually observed in each condition. It will be apparent that a large number of *S*s were suspicious before, or became suspicious during the course of, the experiment. *S*s were classified as suspicious if they acknowledged any doubts about the status of the *SP* during the post-experimental group discussion led by *E*. This is a very stringent criterion of suspicion, since some *S*s were probably influenced by statements of suspicion made by others in their group, or their suspicions may have reached a threshold of awareness during the tedium of filling out the final questionnaires. Nevertheless, for most of the analyses to be presented below, all suspicious *S*s by this criterion were excluded. The fact that 29% of the *S*s expressed some degree of suspicion clearly raises some questions about the representativeness of the remaining sample. More will be said about this later, but the main reason for suspicion is easy to locate: the *SP* was forced by the conditions of the experiment to respond with very vague generalities, and to avoid any reference to the content of the arguments he had presumably received. This unavoidable circumstance combined with the surprise and incredibility of the actual derogation of *S*, made it difficult for some *S*s to accept the *SP* as a naive fourth *S*. It should be noted that eight *S*s were discarded because of failure to comply with the instructions (e.g., they sent unpreferred arguments to *SP*), or because they felt their stand on Big-Time Athletics had been misidentified.

RESULTS

Success of Experimental Manipulations

Response to the SP. The script for the tape recorded response to the *S*'s arguments was constructed with the purpose of instigating the *S*s to hostility. Almost without exception, *S*s did in fact change their ratings in a negative direction. Pooling the change scores of all *S*s

TABLE 1

EXPERIMENTAL DESIGN AND DISTRIBUTION OF *S*S

	Group Support					No Support				
	Pro BTA		Anti BTA		TOT	Pro BTA		Anti BTA		TOT
Rating Form Sequence:	I-II	II-I	I-II	II-I		I-II	II-I	I-II	II-I	
Good *S*s	9	6	6	8	29	6	6	8	7	27
Suspicious[a]	7	5	0	2	14[b]	3	0	4	5	12[c]
Total	16	11	6	10	43	9	6	12	12	39

[a] Although suspicious *S*s appear to be unevenly distributed between conditions, a chi square test reveals that this distribution does not, in fact, depart from chance.

[b] None of these *S*s spontaneously expressed suspicion in the free-response impression sketch.

[c] Five of these *S*s did spontaneously express suspicion.

TABLE 2

MEAN BEFORE, AFTER AND CHANGE SCORES, SUMMARY RATING EVALUATIONS

	Group Support					No Support				
	Pro BTA		Anti BTA		Total	Pro BTA		Anti BTA		Total
Form	I	II	I	II		I	II	I	II	
N	9	6	6	8	29	6	6	8	7	27
Before[a]	3.67	3.92	4.22	4.09	*3.95*	4.63	4.00	4.41	4.26	*4.34*
After[a]	2.81	3.03	3.30	3.35	*3.11*	3.78	3.40	3.79	3.96	*3.75*
Change[b]	.86	.89	.92	.74	*.84*	.85	.60	.62	.30	*.59*

[a] The higher the score, the more favorable the overall evaluation.

[b] The higher the score, the greater the change in the negative direction.

and comparing the mean against the null expectation of zero change results in a t value of 16.4, $p < .001$, leaving little doubt that the *SP*'s recorded comments had their intended effect.

Perception of social support. A crucial condition for an adequate test of the major hypothesis is that the degree of support for the *S*'s arguments, conveyed to him through an exchange of bogus ballots, should be actually perceived as high or low in line with the degree of agreement indicated. Suffice it to say that *S*s in the two conditions are strikingly different from each other ($p < .001$) in amount of agreement perceived (on the recognition–recall measure).

It is possible, of course, that agreement perceived has no implications beyond the purely cognitive ability to recall the bogus ballots distributed. To what extent does each *S* actually infer that the group is behind him, and accepts him as spokesman? Does this inference vary with support treatment? On the postexperimental questionnaire the *S*s were asked to predict whether the other two members of the group would like to have them as "communicator" again. By classifying the responses as yes or no, a chi square test was computed indicating a clear association between support condition and perceived acceptance ($\chi^2 = 5.22$, $df = 1$, $p < .025$). Thus, *S*s in the No Support condition felt that the group was unhappy with their performance and would not be likely to support them in the same role again.

The Effects of Variation in Group Support

Summary rating evaluation. The main hypothesis of the present study was that *S*s in the Group Support condition would show a greater increase in negative evaluation of the *SP* than would *S*s in the No Support condition. This prediction was expected to hold regardless of variation in such control variables as attitude toward athletics and rating form used. Before examining the evidence concerning this main hypothesis, it may be pointed out that neither of these control variables was seen to have any effects of significance in analyzing the main rating data. For presentational simplicity, therefore, no further reference will be made to these variables unless they are specifically relevant to a particular point of difference.

The data most clearly relevant to the main hypothesis are presented in Table 2 and analyzed in Table 3. Here it may be seen that the difference between Group Support and No Support conditions is consistently in the predicted direction for Before, After, and Change scores. The difference between After scores of the two treatment groups is highly significant, the difference between Before scores is significant at the 5% level, and the crucial difference between Change scores does not approach significance.

Interpretation of this pattern of findings is complicated by the quite unexpected difference in Before scores. This result requires explanation in and of itself, because *S*s were randomly assigned to the two treatment groups and at the time the Before measure was taken, *S*s had not been exposed to any differential treatment. The apparent explanation is a curious and unusual one: *S*s in the Group Support condition who initially liked the *SP* tended to become suspicious that he was not responding spontaneously and were later discarded as *S*s; *S*s in the no support condition

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE[a] RESULTS
(Summary rating evaluations: Before, After, and Change scores)

Source	df	Before		After		Change	
		MS	F	MS	F	MS	F
Support (A)	1	.25	4.17[b]	.75	9.38[c]	.12	2.4
Attitude (B)	1	.06		.24		.05	
Form (C)	1	.05		.00		.05	
A × B	1	.07		.00		.02	
A × C	1	.10		.03		.01	
B × C	1	.00		.02		.01	
A × B × C	1	.09		.06		.01	
Error	48	.06		.08		.05	

[a] Because of the unequal cell frequencies, the approximation technique presented by Snedecor (1946) was used.

[b] $p < .05$.

[c] $p < .005$.

who initially disliked the *SP* also tended to become suspicious. The net result of this interaction between support treatment and initial impression, as it affects arousing suspicion, is that usable, nonsuspicious *S*s have significantly different Before scores. The causes of this differential suspicion will be discussed below, but the consequence of the systematic variation in Before scores is the major concern in the present context. It is difficult to interpret the Change scores at face value when the group which is predicted to show the greater decline in rating starts out significantly lower. Both the "floor" effect and the tendency for regression toward the mean would tend to suppress the Change scores in the group support treatment, where the greatest change is predicted. Such a regression effect would be especially likely, given the moderate reliability of the Change scores.

There turns out to be no simple or single way of handling this problem. Two bits of evidence do tend to clarify the picture, however, in the direction of partial support for the major hypothesis of difference between Group Support and No Support conditions. First of all, if the Before scores for the two main treatment conditions are matched by discarding 11 *S*s, a *t* test for matched groups shows the Group Support and No Support conditions to be significantly different ($p < .05$, one-tailed test). This is, of course, a rather wasteful procedure and one which limits the ease of generalization to a broader population of college students.

Secondly, looking at the regression of After scores on Before scores, the coefficient for the No Support *S*s ($b = 1.09$) is significantly greater than the coefficient for the Group Support *S*s ($b = .42$), the probability value of the difference being less than .01. Examination of the scatter plots for the two conditions indicates that this difference is mainly a function of the initially favorable *S*s, who show much more marked decline in the Group Support than in the No Support condition.

Support for this last notion appears when the Before scores for the two treatments are split at the overall median. The difference between Before score means for the two high subgroups is not significant, and these *S*s, who are initially favorable to the *SP*, show significantly greater negative change ($p < .01$) in the Group Support condition than in the No support condition.

With the low subgroups, *S*s in the Group Support condition had significantly lower before scores than those in the Support condition ($p < .01$) but not significantly greater Change scores. Thus, the problem of interpretation concerns only those *S*s who were initially unfavorable; the results for the initially favorable *S*s are straightforward.

To summarize, then, the hypothesis that evaluative ratings will become more negative in the Group Support than in the No Support condition receives only qualified support. A more accurate statement of the general hypothesis would be: those *S*s who show an initially favorable impression of a person who later derogates them for their action will show a greater decline in evaluation when their actions have received normative support than when they have not.

Free response descriptions. In addition to the summary evaluation provided by averaging the item scores on the rating scale, it is important to consider any support for the main hypothesis derived from analysis of the impression sketches contributed by each *S* immediately after the derogation of his arguments. Although the sketches were rich in variety, most of the *S*s seemed concerned with "explaining" the *SP*'s behavior, either by referring to his rigidity, stupidity, and inflexibility, or to his honesty and sense of independence.

Before discussing the effects of the group support variable on these free response descriptions, it is necessary to consider once again the related problems of sampling and suspicion. Since the impression sketches were only written once, after the *S*s had heard both parts of the tape recording, it is obviously impossible to obtain a measure of change from a "before score" baseline. But the significant difference in Before score on the ratings of nonsuspicious *S*s has already been noted, thus tending to confound any differences in the same direction which might be observed on the "after" impression sketches. As a way of handling this problem, the sample was reconstituted for the purpose of this particular analysis. From the total pool of *S*s, only those were discarded who showed any signs of suspicion in the impression sketch itself, although some of them later reported varying degrees of suspicion under questioning. The remaining sample contained 43 *S*s in the Group Support condition and 34 *S*s in the No Support condition. The primary reason for extending the sample in this way was to assure greater comparability between the Before scores (rating scale summaries) of the Group Support and No Support *S*s. This purpose was accomplished since the difference in Before scores between the extended subsamples did not approach significance. We are presumably free, then, to interpret any difference between support conditions on the impression sketches as largely the result of experimental manipulations rather than sampling variation.

Each of the impression sketches was placed by two coders in one of five carefully described rating categories arrayed on the dimension of overall favorability. As evidence of the reliability of coding, 58% of the placements were identical, and 98% of the placements fell in identical or adjacent categories. In cases of disagreement, the final score assigned was the average of the two ratings.

The final distribution of ratings was quite bimodal and skewed, with more of the impression sketches being clearly negative in tone. For this reason, the scores were split as near to the median as possible and the resulting two by two table (More Favorable–Less Favorable, Group Support–No Support) was submitted to a chi square test. The obtained value of 11.97 (corrected for continuity, $df = 1$) s significant at the .001 level, indicating a strong association between support condition and the expression of hostility—those in the Group Support condition express more hostility in their impression sketches than those in the No Support condition.

In general, the Group Support *S*s were more likely than the No Support *S*s to refer to the *SP* as prejudiced, ignorant, egotistical, indifferent, or confused. The No Support *S*s, on the other hand, were more likely to refer to the *SP* as likeable, rational, and honest.

Publicity of Ratings and Group Support

As indicated above, a comparison of the change in summary rating evaluations for Group Support and No Support *S*s provides qualified support for the main hypothesis. That is, the Group Support *S*s show greater negative change than do the No Support *S*s. To what extent does this reflect a private change in judgment, and to what extent are we dealing with differences in the inhibition versus facilitation of the *expression* of hostility? It might be argued that at least some of the *S*s held the vague expectation that they would have to justify their ratings to their fellow group members. If this were the case, then these *S*s would feel more safety in expressing their hostility on the rating scale in the Group Support than in the No Support condition.

No one has yet discovered a simple way of measuring the difference between private feelings and public expression of these feelings. Some form of expression is inevitably bound up with the measurement of attitudes or impressions. However, some partial evidence relevant to this point is provided in the design of the present experiment. It will be recalled that the *S*s filled out two forms of the rating scale after the derogation. First they rated the *SP* on one 10-item form under instructions which stressed the anonymity of the ratings and the fact that they would only be seen by the *E*. This "private" After scale was always the same form as the Before scale. Then they were asked to fill out a different 10-item scale under the explicit instructions that each member's ratings would serve as the basis for a three-person group discussion, a discussion which would lead eventually to a common rating of the *SP*. If the observed changes in rating reported above were entirely a function of inhibition or facilitation of reporting one's hostility, there should be no difference between private and public ratings. If social inhibition and facilitation exercise independent effects, we should expect the difference between Group Support and no support conditions to be even greater on the public ratings.

There are several kinds of analysis that would be appropriate in answering this question, but perhaps the most straightforward involves tabulating the number of *S*s who become more negative and the number of *S*s who become less negative going from the private to public rating form in the Support and No Support conditions. By omitting the few *S*s who did not change in either direction, the resulting two by two table shows that 14 out of 25 Group Support *S*s were more negative in their public ratings, whereas only 6 out of 25 No Support *S*s became more negative and 19 more positive. This difference is significant by chi square test at the .05 level. In other words, when faced with the prospect of publicizing and defending their ratings to the other two group members, the Group Support *S*s became somewhat more hostile on the average. When faced with the same prospect, the No Support *S*s became considerably less hostile on the average.

The interpretation of this finding is complicated by the fact that almost all of this difference is attributable to the group of *S*s rating the *SP* privately on Form II and publicly on Form I. Those *S*s going from Form I to Form II show a slight and nonsignificant change in the same direction. Since there are no differences between Form I and Form II means for any other comparison (Before ratings, Change ratings, private After ratings) it is difficult to know what to make of this variation. However, since the overall effect is significant when both forms are pooled (both by chi square and *t* test) it seems reasonable to state the guarded conclusion that there are independent inhibitory effects stemming from the anticipation of group confrontation in the No Support condition and (to a lesser extent) facilitating effects in the Group Support condition. This does not mean that the private ratings were entirely unaffected by such factors, but it does suggest that they largely reflect the *S*'s internalized impression of the *SP*.

Suspicion as an Alternative to Changes in Evaluation

As noted above, there was no reliable tendency for any of the *S*s to reject the group, devaluate the arguments, or misperceive the degree of support in order to maintain their initial impression of the *SP*. It is possible, however, to consider suspicion as one form of alternative reaction to the experimental situation. As already reported, *S*s in the Group Support condition who initially liked the *SP* tended to become suspicious, whereas, those in the No Support condition who initially *disliked* the *SP* tended to become suspicious. What is the psychological significance of becoming suspicious in the present context? Presumably, *S*s in a psychological experiment become suspicious because (*a*) the situation is poorly rationalized by the *E*, there are incongruous features which destroy credibility, etc., or (*b*) *S* desperately wants to believe that the situation is not as described. In most cases of suspicion these two factors probably reinforce each other—especially when something unpleasant or stressful is done to the *S*. In the present experiment two kinds of *S*s became suspicious. Let us examine each of these in turn.

Considering first those *S*s in the Group Support condition who initially liked the *SP*, it may simply be hard for them to believe that a person for whom they have esteem and respect is responding with such antagonism to so little provocation. The incredulousness of these *S*s may be heightened by their awareness that two other *S*s have checked almost the identical five arguments which were sent. We may speculate, then, that these *S*s became suspicious primarily because of the cognitive–affective incongruity of the experimental situation.

As for those *S*s in the No Support condition who dislike the *SP* initially, the unpleasantness of their position should be noted. They dislike the *SP* and the *SP* appears to dislike them. And yet, the other two group members have indicated that they would have sent different arguments had either of them been communicators. In this situation of unrelieved antagonism and lack of support, it may be that the *S* is highly motivated to perceive any cues of artificiality or incongruity.

The fact that different kinds of people may have become suspicious in the two conditions raises the issue of confounding in the rating results. However, including most of the suspicious *S*s (as in the free response data) results in even greater differences in the predicted direction. Furthermore, there is no a priori reason for assuming that suspicion would consistently bias the results to favor confirmation of the hypothesis.

DISCUSSION

The present research grew out of a general concern with the conditions of interaction which affect the perception and evaluation of a stimulus person. It is to be noted that the behavior of the person being judged was constant and identical from *S* to *S*. The meaning of this behavior, however, was shown to vary with the interaction context from which it stemmed. In the experiment, a stimulus person responded in an abusive, antagonistic fashion to a series of arguments chosen by an experimental *S*. We would naturally expect the *S* to modify his impression of the *SP* in a negative direction, and we have seen that this is indeed the case. But the intensity or degree of negative feelings expressed depends on the implied value *S* attaches to his own eliciting behavior. The more *S* is convinced that his arguments are reasonable, just, and valid, the more righteous and therefore intense we would expect his hostility to become. In the realm of communication about controversial issues, the degree of consensual support for an argument is an important source of its perceived validity (see, for example, Festinger, 1954). The presence or absence of such social support was the main experimental variation in the present study. The results show that *S*s who perceive support for their behavior respond with greater hostility than those who perceive no such support. As predicted, this differential hostility was reflected in the *S*s' verbal impressions of the derogating *SP* and in their attempts to characterize the *SP* on a rating scale (though with the latter measure those who initially disliked the *SP* showed no such differential change).

The present results fit nicely into the discussion by Jones and Thibaut (1958) of the importance of interaction context in person perception, and the method for demonstrating these contextual effects is basically similar to that employed by Jones and Daugherty (1959) and Jones, Hester, Farina and Davis (1959). However, the obtained findings can be accommodated by a number of theoretical positions and seem not to embarrass any others. It is possible, for example, to interpret the results by a simple extension of Heider's (1958) notions of cognitive balance in the P-O-X unit. The individual *S* (*P*) is led to dislike or devalue *SP* (*O*) because they disagree about *S*'s verbal product (X). Following Cartwright and Harary (1956), this particular *P-O-X* unit can be represented as an *s*-digraph in which *P* dislikes *O*, *O* dislikes *P*, *P* likes *X*, and *O* dislikes *X*. Given the fact that *P* likes *X* (*S* values his own chosen arguments) and that *O* dislikes *P* and *X* (as revealed by *SP*'s derogation of both the arguments and the person who thought them up), *P* must dislike *O* for the structure to be balanced. In order to handle the present experimental results, however, it is necessary to introduce a further quantitative assumption: the magnitude of *P*'s liking for *X* will determine the magnitude of his *dis*like for *O*.

In the context of frustration-aggression theory, the present results might be viewed in terms of the degree to which *S*'s hostility toward *SP* is inhibited or facilitated by the Group Support manipulation. *S*s in the Group Support condition may be compared with the members of a cohesive in-group who are secure in the conviction that their beliefs and behavior are not only meaningful but morally right. A considerable amount of anecdotal evidence on witch hunts, lynchings, etc., could be cited to make the point that hostility toward out-group members can be vastly intensified when justified as serving a righteous moral passion. This could be true both because the threat to a highly moral position is viewed as more "frustrating" than the threat to a position more neutral in morality, and because the in-group provides direct positive sanctions for the expression of aggression in the service of a moral cause. The present results provide circumstantial evidence for both of these effects in the laboratory context. There is a difference between the private and the public ratings, the latter seeming to facilitate hostility in the Group Support and inhibit it in the No Support condition. However, the fact that the private ratings show the same difference to a significant if lesser degree, indicates that the present results are not entirely the effect of the group's power to release or suppress aggression in its members. At least we may speculate that this power becomes internalized as the individual responds to others in terms of the normative support inherent in his position.

SUMMARY

The present experiment was concerned with variations in person perception as a function of the perceiver's role in instigating a stimulus person to hostility. While the stimulus person's hostile remarks were always constant across *S*s, the perceiver was led to believe that his instigating behavior either was or was not sanctioned by a group of fellow *S*s. This variation in group support was accomplished by manipulating the extent to which the other group members agreed with a set of arguments transmitted by the perceiver to the stimulus person.

The major findings were generally in line with the hypotheses in that:

1. There was a tendency for the *S*s' evaluative ratings of the stimulus person to become more negative when they perceived that they had group support for their communication (which prompted the derogation by the stimulus person) than when they perceived no such support. This difference was significant when consideration was restricted to those *S*s in each condition who were initially favorably disposed toward the stimulus person.

2. In postderogation free-response "personality sketches," Group Support *S*s evidenced significantly greater hostility toward the stimulus person than did the No Support *S*s.

3. When anticipating the publication of their judgments to the other members of their group, the Group Support *S*s became somewhat more hostile, while the No Support *S*s became considerably less hostile.

In discussing these findings, the importance of considering the interaction context in studying person perception was stressed. The results were also discussed in terms of cognitive balance and frustration-aggression theories.

REFERENCES

CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, **63**, 1956, 277–293.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, **7**, 1954, 117–140.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958, 151–178.

JONES, E. E., & DAUGHERTY, B. N. Political orientation and the perceptual effects of an anticipated interaction. *J. abnorm. soc. Psychol.*, **59**, 1959, 340–349.

JONES, E. E., HESTER, S. L., FARINA, A., & DAVIS, K. E. Reactions to unfavorable personal evaluation as a function of the evaluator's perceived adjustment. *J. abnorm. soc. Psychol.*, **59**, 1959, 363–370.

(Received August 10, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 82-86

DIFFERENTIAL LEARNING AND FORGETTING AS A FUNCTION OF THE SOCIAL FRAME OF REFERENCE

R. KANUNGO AND J. P. DAS
Ravenshaw College, India

RESEARCH on the relationship between the affective tone of material to be learned and retention does not permit any definite conclusion. There have been several obstacles. First, it is very difficult to isolate experimentally the dimension of affectivity from other variables affecting the learning and forgetting processes. Conflicting results and interpretations have also arisen from ambiguities concerning the nature of the feeling tone of the materials used in experiments and the way the materials actually affect the subjects studied.

The usual procedure to test the influence of affective tone on learning and retention is to choose a number of items that have been rated as pleasant, unpleasant, and neutral. These items are learned by the subject and subsequently they are recalled, relearned, or recognized after a specified time interval. In general, the results of various experiments using their procedure go to show that retention decreases in the order of Pleasant-Unpleasant-Neutral. Other experiments using either free recall of past experience or the technique of pairing affective materials with neutral material have yielded conflicting results (Osgood, 1958, p. 571-572).

Seeking to overcome such disagreement, Edwards (cf., Underwood, 1949, p. 523) restated the problem of differential forgetting of affective material not in terms of affect, but "in terms of whether or not the experience or material learned conforms with the individual's frame of reference," (p. 523) that is, his system of evaluation including his values and attitudes.

Osgood (1958) distinguishes between two facets of the problem, one dealing with materials having felt affect and the other dealing with materials having judged affect. Edward's explanation in terms of frames of reference seems most appropriate to retention and forgetting of material having judged affect on "an intellectual or esthetic level." From this point of view, the learning and retention of material judged to have affective tone is influenced by the person's social frame of reference rather than by the feeling tone of the material itself. The present investigation is directed to explore the implications of this explanation of retention of affective materials. *Ss* were selected from two caste groups in India. The materials to be memorized consisted of favorable and unfavorable stereotypes concerning each of the caste groups, and also concerning another plausible but fictitious caste group.

The following hypotheses were formulated:

1. No significant difference should be obtained between learning and retention of favorable and unfavorable adjectives for the nonexistent caste.
2. Regardless of the judged affect, in general, stereotypes concerning one's own group should be better learned and remembered than the stereotypes concerning the other group.
3. Favorable stereotypes for one's own group and unfavorable stereotypes for the other group should be better recalled than favorable stereotypes for the other group and unfavorable stereotypes for one's own group.
4. The effects of Hypothesis 3 should be shown more clearly in delayed recall than in immediate recall.

METHOD

Subjects

Twenty-five Brahmin and 25 Karan undergraduate students served as *Ss*.[1]

Materials

A list of 20 favorable and 20 unfavorable adjectives (control list) was chosen. These were labelled as stereotypes of a fictitious caste group named "Talami." Another list (test list) consisted of two sets of stereotypes, each set consisting of 20 adjectives. One set included 10 favorable and 10 unfavorable adjectives often ascribed to the Brahmin caste, and the other set included 10 favorable and 10 unfavorable adjectives similarly ascribed to the Karan caste.[2] All adjectives were in Oriya language, the mother tongue of the *S*s. These favorable and unfavorable stereotypes relating to the two castes were taken from an earlier study.[3]

---

[1] Brahmin, the priest's caste, and Karan, the caste of the clerks and ministers to the King in old India, are the two top classes in the part of the country in which the study was conducted. Each of the two castes considers itself as the aggressive competitor of the other in social, educational, occupational, and professional spheres of life. A certain amount of antagonism exists between them, which has been taken advantage of in the present study.

Procedure

The design of the experiment was determined by the need to provide different frames of reference to *S*s who were engaged in learning and recalling materials that had different judged affect for them. Therefore the study had two stages. At the initial stage the *S*s were presented the list of 40 adjectives labelled as stereotypes of the fictitious caste "Talami" with the following instructions:

> In a particular place in India there is a group of people known as Talami. Authoritative reports about them indicate that they possess the following attributes which you shall hear one by one. Please try to be attentive to what is being said about the group and try to remember them to the best of your ability. This is a test of your memory. Now listen.

The list of 40 adjectives was then presented to the *S*s by tape recorder. Word order was randomized using a table of random numbers. After eight consecutive presentations of the list *S*s were asked to write down as many adjectives of the list as they could. On the next day about 24 hours after the original learning *S*s were again asked to recall the learned material. To avoid rehearsal within the intervening period care was taken not to give any hint about the delayed recall.

On the basis of their recall scores, both immediate and delayed, 15 Brahmin and 15 Karan *S*s were chosen so as to form two equivalent groups for the next stage of the experiment.

The second stage of the investigation began two months after the initial test. This time the test list of stereotypes was given, with the following instructions:

> Earlier you were tested for your memory ability. To know whether, in the meantime, you have deteriorated or improved in your memory ability, we shall have a fresh test. Last time the list consisted of stereotypes towards the Talami group whom you never knew before. But this time you will be given two sets of stereotypes, one relating to Brahmins and another relating to Karans whom you know quite well. Try to remember what is said about the two castes separately. As this time we just want to check our earlier results, you would not be given a delayed recall test.

TABLE 1

MATCHING OF GROUPS ON RECALL SCORES, INITIAL TEST

(In each group, $N = 15$)

Recall	Caste	*M*	*SD*	*t*	*P*
Immediate	Karan	25.66	3.52	.01	NS
	Brahmin	25.62	4.82		
Delayed	Karan	22.13	3.68	.28	NS
	Brahmin	21.66	4.76		

TABLE 2

RECALL SCORES OF BRAHMINS AND KARANS

(In each group, $N = 15$)

Caste	Nature of Stereotypes	Immediate Recall *M*	Immediate Recall *SD*	Delayed Recall *M*	Delayed Recall *SD*
Brahmin	Favorable stereotypes towards themselves	5.33	1.63	3.67	.45
	Unfavorable stereotypes towards themselves	5.27	1.97	4.73	2.13
	Favorable stereotypes towards others	3.6	2.3	3.13	2.24
	Unfavorable stereotypes towards others	4.67	1.35	4.07	1.77
Karan	Favorable stereotypes towards themselves	5.8	1.55	4.40	1.86
	Unfavorable stereotypes towards themselves	5.93	1.70	3.67	1.48
	Favorable stereotypes towards others	4.6	1.96	3.27	1.65
	Unfavorable stereotypes towards others	6.2	1.30	5.53	1.51

Then the two sets of stereotypes of the test list, one relating to Brahmins and the other to Karans, were presented by tape recorder. Each set contained equal number of favorable and unfavorable stereotypes in randomized order. The two sets were presented in A-B-B-A order: stereotypes relating to Brahmins, then stereotypes relating to Karans, stereotypes relating to Karans, stereotypes relating to Brahmins. This sequence was repeated consecutively for four times. The instructions were framed to minimize rehearsal that might have occurred had the *S*s anticipated measurement of delayed recall. After about 24 hr., however, measures of delayed recall were taken.

RESULTS AND DISCUSSION

An examination of Table 1 reveals that an attempt to select equated groups in the Initial test was successful.

The first hypothesis predicted no significant difference in learning and recall of favorable as against the unfavorable adjectives ascribed to the fictitious caste. This was borne out by a *t* test.

---

[2] The lists of stereotypes comprising the initial and final tests have been deposited with the American Documentation Institute. Order Document No. 6316, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm, or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] Unpublished research on social tension, by R. Rath (1958).

The results of the main experiment are presented in Table 2. An analysis of variance was performed on the scores of the main experiment and the results are given in Table 3. A 2^n factorial design was adopted (Edwards, 1950) since there were four main variables each dichotomously divisible. All the main effects and only one interaction were significant. The second hypothesis, which predicted a significant difference between recalling adjectives relating to one's own caste vs. those relating to the other caste, is supported by the significant F for main Effect 4. Each group appears to be more interested in what is being said about itself, irrespective of whether the adjectives are favorable or unfavorable.

TABLE 3

ANALYSIS OF VARIANCE OF RECALL SCORES

Source	Sum of Squares	*df*
Immediate & delayed recall	74.8167***	1
Castes	22.8167**	1
Favorable & unfavorable	36.8167***	1
Favorable and unfavorable towards themselves and favorable and unfavorable towards others	13.0667*	1
1 × 2	5.400	1
1 × 3	0.600	1
1 × 4	7.3500	1
2 × 3	0.0667	1
2 × 4	10.4167	1
3 × 4	28.0167**	1
1 × 2 × 3	1.3500	1
1 × 2 × 4	0.2677	1
1 × 3 × 4	0.0667	1
2 × 3 × 4	11.2667	1
1 × 2 × 3 × 4	7.3500	1
Residual	720.054	224
Total	939.734	239
Residual msv.	3.2145	

Note.—F significant:
* $P < .05$
** $P < .01$
*** $P < .001$

The only significant interaction may be taken to support the third hypothesis which predicted a differential recall of favorable and unfavorable adjectives relating to ones own and to the other group. But does this overall significant interaction mean that for each caste and both for immediate and delayed recall, favorable adjectives for one's own group and unfavorable ones for the other group are better recalled? Further analysis presented in Table 4 shows that this is not the case. Contrary to predictions, there was no difference in the immediate recall of unfavorable stereotypes relating to one's own and to the other group. Only in the Karan caste was the predicted difference in favorable stereotypes obtained for both immediate and delayed

TABLE 4

RECALL OF FAVORABLE AND UNFAVORABLE STEREOTYPES AS DIRECTED TOWARDS OWN CASTE OR TOWARDS OTHER CASTE

Caste	Nature of Stereotypes	Immediate Recall					Delayed Recall				
		Mean_D	SD_D	SE_M	*df*	*t*	Mean_D	SD_D	SE_M	*df*	*t*
Karan ($N = 15$)	Favorable stereotypes towards themselves > towards others	1.2	1.82	0.47	14	2.35*	1.13	1.88	0.48	14	2.35*
	Unfavorable stereotypes towards others > towards themselves	0.26	1.45	0.37	14	0.70	1.86	0.99	0.25	14	7.44***
Brahmin ($N = 15$)	Favorable stereotypes towards themselves > towards others	1.73	1.43	0.37	14	4.67***	0.53	1.27	0.32	14	1.65
	Unfavorable stereotypes towards themselves > towards others	0.60	1.82	0.44	14	1.36	0.66	1.23	0.31	14	2.12

* $P < .05$.
*** $P < .001$.

TABLE 5

RATING OF STEREOTYPES BY 29 JUDGES INTO FIVE CATEGORIES

Frequency of Use	No. of Ratings Favorable Stereotype ($N = 20$)		No. of Ratings Unfavorable Stereotype ($N = 20$)	
Very frequently used	122	230	249	378
Frequently used	108		129	
Neither frequently nor infrequently used	119		75	
Less frequently used	117	231	88	127
Least frequently used	114		39	
Total	580		580	

recall. However, except for the recall of unfavorable stereotypes by Brahmin *S*s, the differences were in the predicted direction.

The fourth hypothesis is only partially upheld by the significant difference in the delayed recall of unfavorable stereotypes by the Karan *S*s.

Besides confirming the hypotheses under study, the results of the analysis of variance pointed up two significant main effects which were not predicted: castes and favorable-unfavorable stereotypes. Although the castes were matched in the initial test, the Brahmin *S*s are found to have a significantly lower recall score than the Karan *S*s in the main experiment. Perhaps the Brahmin group is more heterogeneous, for on the whole, it has larger *SD*s than the Karan group (see Tables 1 and 2).

The significant difference in the recall of favorable and unfavorable stereotypes calls for closer attention. The most plausible explanation as to why unfavorable stereotypes are better recalled may be that *S*s are more familiar with the unfavorable stereotypes, possibly because these are more often used in spoken language. In order to test this hypothesis, all the 40 stereotypes used in the main experiment were given to 29 judges in a predetermined random order for rating. The judges, who were college teachers, rated each of the sterotypes on a five-point scale representing the extent of use of these words in common spoken language. The results are presented in Table 5 and can be seen to support the hypothesis that unfavorable stereotypes are more familiar. Leaving out the middle category, the unfavorable stereotypes receive 378 ratings as frequent as against 127 infrequent ratings; but the comparable figures for favorable stereotypes are 230 and 231. Again, the unfavorable stereotypes get 249 ratings as "very frequently used" whereas the corresponding figure for favorable stereotypes is 122. It was indeed a surprising revelation to the authors that in their language unfavorable stereotypes like traitor, swindler, and bribe-taker find greater favor than such favorable ones like cultured, hospitable, or artistic.

SUMMARY

The present study was aimed at demonstrating how learning and recall become selective because of the relevant frames of reference. Two groups of 15 *S*s each, were selected from two competing caste groups, Brahmins and Karans, and were required to learn and recall 20 favorable and 20 unfavorable familiar stereotypes concerning themselves. The groups were matched on the basis of their recall scores in a preliminary test.

It was hypothesized that (*a*) there should be no difference in learning and retention of favorable and unfavorable stereotypes for a fictitious caste, (*b*) stereotypes relating to one's own caste should be better learned and remembered than those relating to the other caste, (*c*) favorable stereotypes for one's own caste and unfavorable stereotypes for the other caste should be better learned and recalled than favorable stereotypes for the other caste and unfavorable stereotypes for ones own caste, and (*d*) this will be shown more clearly in delayed recall than in immediate recall.

A 2^n factorial analysis of variance showed that Hypotheses 1, 2, and 3, were valid, the last one only generally. Hypothesis 4 was only partially supported.

Unfavorable stereotypes were better recalled than the favorable ones. A supplementary test suggested that this was so because the unfavorable stereotypes were more often used

in the language of the *S*s than the favorable ones.

REFERENCES

EDWARDS, A. L. *Experimental designs in psychological research.* New York: Rinehart, 1950.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1958.

UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

(Received July 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 87-92

INTERPERSONAL KNOWLEDGE AND RATED LEADER POTENTIAL[1]

MORRIS SHOWEL
Human Resources Research Office, George Washington University[2]

It is a widely accepted principle of military leadership that a leader should know his men. In psychological terms, this could be translated to mean that a leader should demonstrate "empathy," "social sensitivity," or "social perception." In spite of a considerable amount of research on social perception and the relation between ability in this area to some criterion of effectiveness (Taft, 1955), the validity of the principle has not yet been satisfactorily demonstrated. Moreover, considerable criticism has been levied against the conceptual approaches and the research procedures used to study the problem in the past. (Cronbach, 1955; Gage & Cronbach, 1955.)

This paper presents a somewhat different approach toward measuring social perception and relating it to a criterion of effectiveness. First, the concept of social perception has been broadened to include not only perception of attitudes, opinions, and preferences, but of demographic data as well. Second, rather than dealing with accuracy of social perception of group characteristics, this study deals with accuracy of social perception about the particular individuals comprising the group. Bush (1954) used a similar procedure and found a positive correlation between teacher effectiveness and the amount of knowledge teachers had about their students. Using the procedure just described seems to avoid some of the criticisms raised by Gage and Cronbach (1955) to previous research in this area: the problem of real similarity, of stereotype accuracy, and of securing an independent criterion of effectiveness in interpersonal relations.

METHOD

Subjects and Setting

Subjects for this study were infantry trainees in the process of completing their six-month tour of duty. These trainees comprised two of the five platoons of one training company. Each of the platoons contained four squads, and each squad contained approximately ten nonleader trainees. The company was led by a company commander (officer), whose chief administrative assistant was a first sergeant (noncommissioned officer). Each platoon was led by a platoon leader (officer) who was assisted by a cadre platoon sergeant (noncommissioned officer). In addition, each platoon had a trainee platoon sergeant and a trainee platoon guide who acted in a leadership capacity below the platoon leader and the cadre platoon sergeant. Each squad was led by a trainee squad leader.

Platoons were housed in separate barracks, two squads per floor. Squad leaders lived with their men. The trainee platoon sergeant and the trainee platoon guide lived in separate rooms in their platoon's barracks. The cadre platoon sergeant, platoon leader, first sergeant, and company commander did not live in the barracks.

All of these personnel had served together for a minimum of six weeks, some of the trainees for as long as six months. The intimate nature of barracks living and the extended period of time during which trainees served together provided considerable opportunity for the trainees to get to know one another.

Interpersonal Knowledge Questionnaire

Accuracy of social perception was measured by means of an Interpersonal Knowledge Questionnaire. The information asked for in the questionnaire was selected on the basis of three criteria: (*a*) the information would be useful for a leader to know; (*b*) it could easily be verified; (*c*) it could be secured via a questionnaire or brief interview. The questionnaire included a report on self, on the men in the squad, on the trainee platoon sergeant and guide, and on the cadre (cadre platoon sergeant, platoon leader, first sergeant, and company commander).

Report on self. All trainees answered the following questions about themselves: (*a*) first name; (*b*) last name; (*c*) whether or not during the past week they had been on KP; (*d*) whether or not during the past week they had been on sick call or in the hospital; (*e*) whether or not during the past week they had been on pass or leave; (*f*) their rifle qualification score; (*g*) the amount of schooling they had completed; (*h*) their main job or activity before entering the Army; (*i*) their principal hobby or interest; (*j*) their ambition for a future civilian career.

---

[1] The opinions and conclusions contained herein are those of the author and are not to be construed as representing those of George Washington University or the Department of the Army.

[2] Located at United States Army Leadership Human Research Unit, P.O. Box 787, Presidio of Monterey, California.

Report on trainee squad leaders and nonleaders. Trainee squadleaders and trainee nonleaders were given the last names of the men in their squad, including the squad leader, and then asked to answer all of the other questions about these people. Reports on trainee squad leaders and nonleaders were then compared with the self-reports to determine the amount of interpersonal knowledge possessed. As a result of the relatively objective nature of the data, all scoring was done by one person. Items #8, #9, and #10 required the most interpretation on the part of the scorer. The objective nature of the data also makes it evident that relatively little extrapolation was required of the *S* when reporting on others. Each correct answer was awarded one point. Neither blanks nor incorrect answers were counted. Since the squads differed somewhat in size, a correction for these differences was handled by summating the points received by each man, dividing this total by the number of men in his squad, and then multiplying it by 10, to give a score with a ceiling of 90 points.

Report on trainee sergeants and guides. Trainee squad leaders and trainee nonleaders were asked to answer Questions 1, 2, and 4 through 10 about their trainee platoon sergeant and trainee platoon guide. (Trainee guides and sergeants do not serve as KPs.) Scoring was done in the same manner, for a possible 18 points.

Report on cadre. Trainee squad leaders and trainee nonleaders were asked to write: (*a*) the first name; (*b*) the last name; (*c*) the exact rank of their company commander, first sergeant, platoon leader and cadre platoon sergeant.[3]

Reports on cadre were then compared with information secured from these people via interview. Scoring was done in the same manner, for a possible 12 points.

Total knowledge. A total knowledge score was computed by summing the scores received by each *S* on cadre, guide and sergeant, and squad leaders, and nonleaders (corrected for size of squad). Knowledge about trainee squad leaders and nonleaders contributed most heavily to the score.

Rated Leader Potential

Five measures of leader potential were secured.

Trainee squad leaders and nonleaders ratings. Each member of a squad, including the squad leader, ranked the men in his squad, including the squad leader, in terms of the kind of a squad leader the man would make. (At the time of the study, the squad leaders were actually *acting* squad leaders.) Following the ranking, the *S*s assigned each man whom he had ranked a numerical weight. These weights ranged from 1 ("the very best type of squad leader") to 7 ("the worst kind of squad leader"). An individual's trainee squad leader and nonleader rating was the mean of the weights he had been given by his squad mates. His rank order, as such, was not used in the subsequent analysis.

Trainee sergeants and guides ratings. The trainee platoon sergeant and the trainee platoon guide ranked and rated the trainees in their platoon in the same manner. An individual's trainee sergeants and guides rating was the mean of these two ratings.

Cadre rating. The cadre platoon sergeant and the platoon leader ranked and rated the men in their platoon and cadre ratings were computed in the same manner as for trainee sergeants and guides.

Total superiors' ratings. A total superiors' rating was computed by summing the weights assigned by trainee sergeants, trainee guides, platoon sergeant, and platoon leader, and dividing this value by four.

Leader reaction test. Forty-three of the trainees were tested by means of a leader reaction battery (United States Department of the Army, 1958a). Three of the tests in the battery required the *S* to perform as a leader in tactical situations, two of the tests required him to take part in leaderless discussion groups, and one test required him to lead a squad through dismounted drill. The followers in the tactical tests (three different persons for each test) were not acquainted with the *S*. The followers in the two discussion groups and in the dismounted drill test were generally from the same platoon, but not necessarily from the same squad. Each of the six tests lasted approximately 15 min. Ratings on all six tests were made by trained raters who had never seen the *S* prior to the test. A *S*'s leader reaction test score is the weighted average of his scores on all six tests.

Intelligence

In order to control for intelligence, the general-technical aptitude area score of each *S* was secured. The GT score is made up of the verbal intelligence and arithmetic reasoning examination scores received by each trainee when he entered the Army.

Method of Analysis

Because of the prominent and highly visible position occupied by trainee platoon sergeants, platoon guides, and squad leaders, it was decided to remove them from the larger trainee nonleader group before examining the relationship between the knowledge variables, the leader potential variables, and intelligence by means of product-moment and partial correlation coefficients.

Comparisons were also made between the average scores of trainee squad leaders and trainee nonleaders. Significance of differences between means, when the variances and *N*s are unequal, was tested by means of a procedure proposed by Cochran and Cox and cited by Johnson (1949).

RESULTS AND DISCUSSION

Leader-Nonleader Comparisons

As can be seen in Table 1, the average scores of the trainee squad leaders were consistently more favorable than the average scores of the nonleader trainees. It appears that trainee infantrymen acting in the capacity

---

[3] Company commanders, first sergeant, platoon leader, and platoon sergeant are job titles, not rank designations.

TABLE 1

SCORES RECEIVED BY TRAINEE SQUAD LEADERS AND NONLEADERS ON INTERPERSONAL KNOWLEDGE QUESTIONNAIRE, GT, AND LEADER POTENTIAL

	Trainee							
	Squad Leaders			Nonleaders			t	p
	M	N	SD	M	N	SD		
Knowledge Regarding								
1. Cadre–12 points possible	8.8	7	1.46	7.8	78	2.25	2.38	*NS*
2. Trainee sergeants and guides–18 points possible	12.8	7	2.36	8.1	78	3.63	4.52	.01
3. Trainee squad leaders and/or nonleaders–90 points possible	47.0[a]	7	7.88	37.1	78	14.46	2.73	.05
4. Total knowledge–120 points possible	68.7	7	8.90	53.0	78	18.27	3.45	.01
Intelligence								
5. GT examination score	111.2	8	19.86	97.8	78	22.36	1.69	*NS*
Leadership ratings given by								
6. Trainee squad leaders and/or nonleaders (possible range 1 to 7)[b]	2.1[a]	8	.82	3.8	78	.97	5.18	.001
7. Trainee sergeants and guides (possible range 1 to 7)[b]	2.2	8	.83	4.5	78	1.32	6.63	.001
8. Cadre (possible range 1 to 7)[b]	2.5	8	.61	3.9	78	.77	5.64	.001
9. Total leaders (possible range 1 to 7)[b]	2.3	8	.62	4.2	78	.94	7.34	.001
10. Leader reaction test	105.0	4	10.82	76.1	43	16.01	4.29	.02

[a] Trainee squad leaders answered interpersonal questions about trainee nonleaders only and only received ratings from trainee nonleaders.

[b] Lower scores indicate higher ratings.

of squad leader receive higher leader potential ratings, perform better in the leader reaction tests, and have more accurate interpersonal knowledge about their associates than trainee infantrymen not acting in the capacity of squad leader. There is no significant difference between the two groups with regards to intelligence.

At least three explanations are possible for these findings. First, possession of interpersonal knowledge could increase the trainee's ability to deal with others, and lead the trainee's superiors to consider him a potential leader and place him in a leadership position. Second, it is possible that the increased amount of interpersonal knowledge was secured *after* the trainee is in a leadership position. The need to deal with others, both subordinates and superiors, may lead the trainee to acquire the knowledge necessary to do the job. On the other hand, the higher leader potential ratings could be a function of being in an official leadership position. The position of squad leader is conspicuous, and a man could receive a high rating by virtue of occupying and being highly visible in this leadership position. A third possible explanation might be that ratings, position, and knowledge are all symptomatic of another underlying variable.

Nonleader Relationships

In order to control for the highly visible position occupied by the trainee leaders, the remaining discussion is confined to data (Table 2) involving trainees who did not hold a leadership position at the time the study was conducted.

Knowledge. The three independent measures of interpersonal knowledge (Variables 1, 2, and 3) are positively and significantly correlated with one another. Nonleaders who were well informed about the cadre were also well informed about the trainee sergeants and guides, and about the trainee squad leaders and nonleaders. The total knowledge score (Variable 4) shows a moderately strong relation to Variables 1 and 2, and a near perfect relation to Variable 3. It should be remembered that the correlations of the three knowledge scores (Variables 1, 2, and 3) with the total knowledge score (Variable 4) are part-whole correlations and hence spuriously high.

TABLE 2

CORRELATIONS BETWEEN INTERPERSONAL KNOWLEDGE, LEADER POTENTIAL, AND INTELLIGENCE FOR NONLEADER TRAINEES

	1*	2*	3*	4*	5*	6*	7*	8*	9*	10**
Knowledge Concerning										
1. Cadre		.46	.58	.67	.40	.18	.14	.16	.15	.34
2. Trainee sergeants and guides			.58	.71	.44	.21	.31	.15	.27	.25
3. Trainee squad leaders and nonleaders				.98	.50	.32	.43	.15	.36	.37
4. Total knowledge					.53	.32	.42	.17	.36	.38
Intelligence										
5. GT examination score						.33	.27	.21	.28	.44
Leadership ratings given by										
6. Trainee squad leaders and nonleaders[a]							.71	.51	.72	.07
7. Trainee sergeants and guides[a]								.52	.92	.14
8. Cadre[a]									.80	.13
9. Total leaders[a]										.15
10. Leader reaction test										

* *N*-78, .233 significant at .05 level; .320 significant at .01 level.

** *N*-43, .304 significant at .05 level; .393 significant at .01 level.

[a] Negative coefficients involving Variables 6, 7, 8, and 9 have been reversed.

Leader ratings. The three ratings of leader potential (Variables 6, 7, and 8), are positively and significantly correlated with one another. Nonleader trainees rated as having high leader potential by the cadre are also rated as having high leader potential by trainee sergeants and guides and by trainee squad leaders and nonleaders. Some differences might have been expected in view of the background differences between trainees and cadre. The trainees were all six-month soldiers, having enlisted for a six-month tour of duty in preference to being drafted for two years or enlisting for three or more years. The cadre, on the other hand, were more professional military men. Nevertheless, it appears that cadre and trainees used similar standards in evaluating the leader potential of the trainees.

While trainees and cadre appear to look at leadership in a somewhat similar manner, it is nevertheless apparent that the two trainee groups (sergeants and guides, and squad leaders and nonleaders) are more in agreement with one another than are either with the cadre. The difference in background between trainees and cadre appears to be the most reasonable explanation of this pattern.

Leader reaction test. Scores on the leader reaction battery (Variable 10) do not correlate significantly with any of the other ratings of leader potential. This finding suggests that the dimensions of leadership considered by the raters and those measured by the tests are quite different. It is not possible to say with any certainty what dimensions were considered by the raters when they made their ratings. It seems likely however, that the trainee raters took into consideration human relations skills when they made their ratings,[4] since they had lived and worked together for a minimum of six weeks. The performance tests, on the other hand, made few if any demands on human relations skills. The tests were primarily concerned with the trainee's ability to use instantaneously the technical military skills he had been taught. The followers in most cases were men with whom the trainee had not developed any working relationship, and the brief duration of the tests prevented him from developing such a relationship.

Knowledge and leadership. The more knowledge a nonleader trainee had about his trainee squad leader and nonleader colleagues, the higher was the leader potential rating he received from *all* trainees. This relationship cannot be accounted for by the position that these nonleaders occupied, for their position in the squad and platoon, both with regard to visibility and authority, was no different from that of nonleaders with low ratings and low levels of interpersonal knowledge. Again, a number of explanations are possible. First, in the process of securing interpersonal information, the high rate might have given indication of a high degree of personal interest in their co-workers, who in turn may therefore have been more likely to rate them higher in leader potential. Second, having such information might enable a high ratee to deal more effectively with his co-workers. This in turn might lead co-workers to give these trainees higher leader potential ratings. Third, general intelligence, or some other factor, might account for both the high level of interpersonal knowledge and the high leader potential ratings.

---

[4] Another questionnaire (completed by a larger group of *S*s) required a ranking in order of importance of 10 qualities that a squad leader should have. Both cadre and trainees placed the same five qualities in the upper half of the ranking. The most important quality for both groups was "willingness to accept responsibility," but the cadre gave it a mean rank of 2.3 while the trainees gave it a mean rank of 3.4. The second most important quality for the cadre was "to be efficient in his job" (mean rank 3.1). This quality was fifth in importance for trainees (mean rank 4.9). The second most important quality for the trainees was "to understand the men" (mean rank 4.4). This quality was fifth in importance for the cadre (mean rank 5.3). This analysis was carried out by C. W. Peterson.

Knowledge about trainee sergeants and guides correlated positively and significantly with the leadership ratings given by these people. Knowledge about the cadre was not significantly correlated with the leader potential ratings given by either trainees or the cadre.

Having accurate information about co-workers, either trainee or cadre, appears to have little if any bearing upon the leader potential ratings received from the cadre. It appears here that the dimension of leadership tapped by the Interpersonal Knowledge Questionnaire was not considered by the cadre when they made their ratings.

Intelligence. Intelligence, as measured by GT, correlated positively and significantly with all four measures of interpersonal knowledge, and with all but one of the five measures of leader potential. Using the total knowledge score as the measure of interpersonal knowledge, the relation between interpersonal knowledge and the various measures of leader potential, holding intelligence constant, are as follows: with trainee squad leaders and nonleaders ratings, $r = .18$; with trainee sergeants and guides ratings, $r = .34$; with cadre ratings, $r = .07$; and with the leader reaction battery scores, $r = .20$. Only the correlation with trainee sergeants and guides ratings is statistically significant.

It thus appears that a considerable portion of the variation in leader potential originally attributable to interpersonal knowledge may be attributed to intelligence. Yet not all can be so attributed. A moderate relation remains between total interpersonal knowledge (primarily a measure of knowledge about trainee squad leaders and nonleaders) and ratings given by trainee sergeants and guides. This is particularly interesting, for in many respects, the trainee sergeants and guides may have been the most valid raters. They had far greater opportunity than the cadre to observe the day-to-day performance of the nonleader trainees, and their conception of what constitutes the military requirements of leadership were probably more accurate than the conceptions of military leadership held by trainee squad leaders and nonleaders.

The obtained partial correlation of .34 between interpersonal knowledge and trainee sergeants and guides ratings is not particularly high in a practical sense, but the fact that it is significant suggests that there is some empirical support for the generally accepted leadership principle "know your men" (USA, 1958a).

SUMMARY

This study investigated the relationship between the possession of interpersonal knowledge about others and the ratings received as to leader potential. Four measures of interpersonal knowledge and five measures of leader potential were secured. *S*s were two platoons of soldiers completing a six-month tour of duty in the Army.

The data indicate that the more interpersonal knowledge nonleader trainees had, the higher was the leader potential ratings they received from trainee leaders and trainee nonleaders. The correlation drops when intelligence is controlled, but that between total knowledge possessed and ratings received from one group of trainee leaders, trainee sergeants and guides, still remains statistically significant. It is hypothesized that the ratings given by trainee sergeants and guides are more valid than the ratings given by the cadre, by trainee squad leaders, and by trainee nonleaders.

Additional findings were that trainee squad leaders had more interpersonal knowledge and received higher leader potential ratings than trainee nonleaders. Trainee squad leaders and trainee nonleaders did not differ significantly in regard to intelligence.

REFERENCES

BUSH, R. N. *The teacher-pupil relationship.* New York: Prentice-Hall, 1954.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

JOHNSON, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.

TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.

UNITED STATES DEPARTMENT OF THE ARMY. *Military leadership*, FM 22–100. Washington, D. C.: USA, 1958. (b)

UNITED STATES DEPARTMENT OF THE ARMY, Adjutant General's Office. *Manual for administering and scoring the Leadership Performance Battery*, PT 3579. Washington, D. C.: USA, 1958. (a)

(Received July 20, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 93-101

COERCIVE POWER AND FORCES AFFECTING CONFORMITY[1]

JOHN R. P. FRENCH, JR., H. WILLIAM MORRISON,
University of Michigan

AND GEORGE LEVINGER
Bryn Mawr College

WHAT determines the conformity of a subordinate to the orders of his superior? In a more general theory, based on Lewin's constructs, we have distinguished five bases of power that might lead to such conformity: legitimate, coercive, reward, referent, and expert power (French & Raven, 1959). The theory focuses on P, the person being influenced by the other person, O. Two previous experiments demonstrated that the overt behavior of P conforms to the orders of O when this superior has legitimate power over P and when he has coercive power. (Raven & French, 1958a; 1958b). But there was a tendency for P's private opinions to conform less when O used coercion. To account for these findings it seemed necessary to assume that P's behavior was determined both by induced punishment forces and by opposing resisting forces. However, no measures of these intervening forces were obtained, so there was no direct test of the interpretation. The purpose of this experiment is to test directly this theory about the effects of punishment forces and resisting forces on conformity.

The structure of the theory is outlined in Fig. 1, with the independent (experimental) variables at the top of the figure, the dependent variable at the bottom, and the intervening variables in between. The solid arrows indicate the major hypotheses or assumptions, while the dotted arrows indicate hypotheses that may be derived from two other hypotheses. The formal statement of these hypotheses is given in connection with presenting the relevant results, and the numbering of the arrows in Fig. 1 corresponds to the numbers of these hypotheses. Thus, the data for each relation can be found in the section on results under the hypothesis whose number is given in parentheses. The other numbers in Fig. 1 show the derivations of the derived hypotheses.

The central part of the theory states that the amount of conformity of a subordinate to his superior depends on the balance of the punishment force and the resisting force. The punishment force is defined as a tendency toward conformity that is set up by O's act of coercion—either a punishment or a threat of future punishment. A punishment is defined as an intentional act of O which places P in a region of negative valence; an accidental act of O which harms P is not perceived as a punishment. *Threats* of future punishment similarly place P in a region of negative valence, but only to the extent that the threatened punishment is probable; if P does not believe the threat, he does not anticipate punishment. If P perceives that O will probably carry out his intentional threat of punishment for nonconformity, then P's conforming behavior is instrumental to avoiding the punishment. Thus, the threat sets up punishment forces that produce conformity (Hypothesis 1 in Fig. 1). But actual punishment should be effective in producing subsequent conformity only if P expects that the punishment will be continued or repeated (Hypothesis 2).

The strength of the *punishment force* based on an expectation of punishment by O and directed toward compliance with O's influence attempts could not be measured directly in this experiment. Instead we calculated its strength on the assumption (cf., Fig. 1) that it depended on the valence of the punishment and the differential probability of receiving it contingent upon nonconformity vs. conformity. The formula was:

$$\text{Punishment force} = (\text{valence of punishment}) \times (\text{probability of punishment if nonconformity} - \text{probability of punishment if conformity})$$

From the assumption expressed in this formula and from Hypothesis 1 we may derive the prediction that threatened Ss in the coercive experimental treatment will show

---

[1] This experiment was conducted at the Research Center for Group Dynamics under contract with the Office of Naval Research.

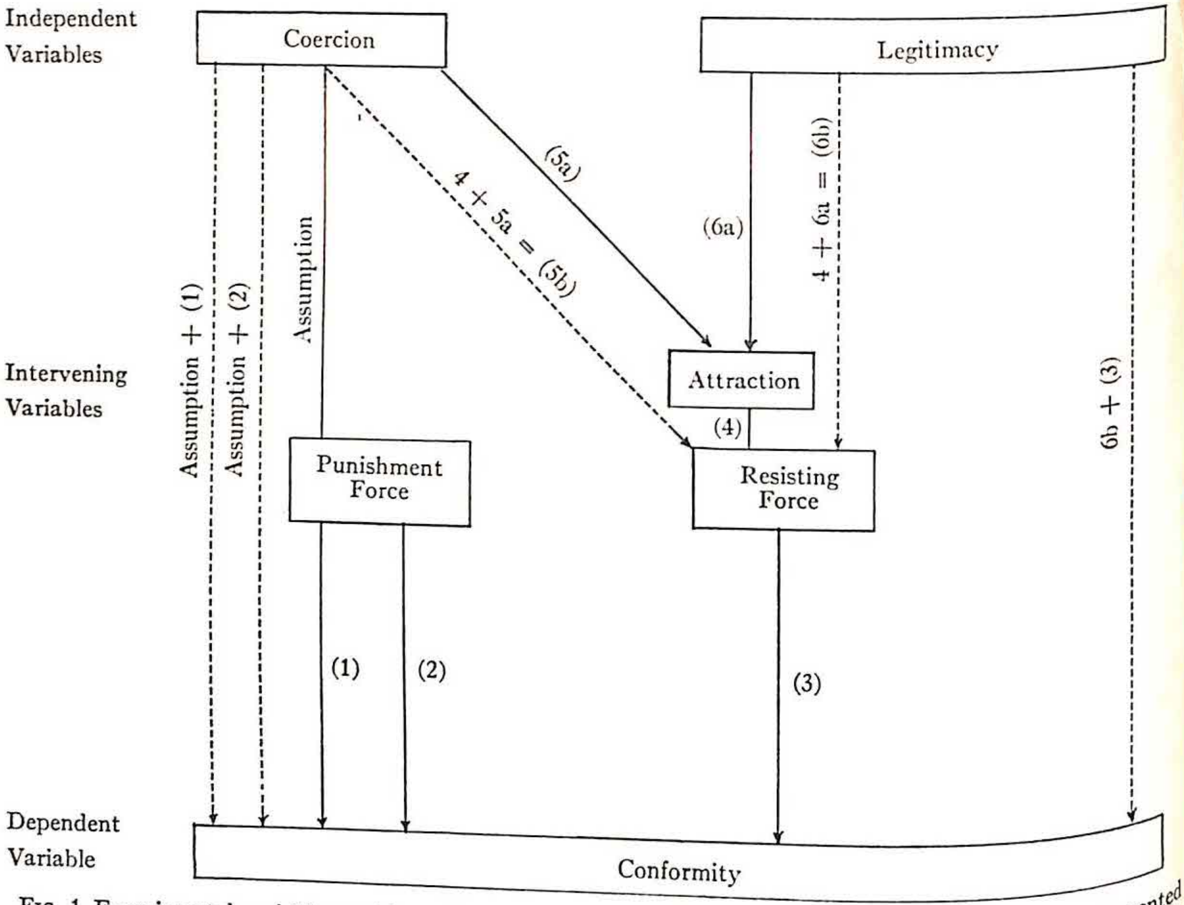

FIG. 1 Experimental variables and hypotheses testing their relationship. (The main hypotheses are represented as solid arrows with their encircled numbers attached; derived hypotheses and supplemental tests are represented as dotted arrows which are labelled so as to show the derivation, with the encircled number indicating where the test can be found in the text.)

more conformity than *S*s in a noncoercive treatment. Similarly, the assumption and Hypothesis 2 leads to the derivation that no such difference will occur after the punishment is over and no further punishment is possible.

A *resisting force* is defined as a force resulting from the same act of *O* that gave rise to the punishment force but in a direction opposite to that of the punishment force. Such resisting forces should reduce conformity (Hypothesis 3).

But what determines the strength of resisting forces? In the previous experiment low attraction to the supervisor was produced by both the supervisor's utilization of coercion (Hypothesis 5a) and by his illegitimate occupancy of the supervisory role (Hypothesis 6a). Therefore, it seemed reasonable to hypothesize that low (or negative) attraction produced resistance (Hypothesis 4) and that attraction mediates the effects of coercion on resisting force (Hypothesis 5b) and the effects of legitimacy on resisting force (Hypothesis 6b). However, it may be that *P*'s values are more important in mediating the effects of legitimacy on resisting forces. Perhaps the more *O*'s coercion is perceived as legitimate or "right," the more would resistance to his coercive influence attempts be incompatible with *P*'s value system regarding *O*'s role and the appropriate use of punishment to gain conformity.[2]

METHOD

Design and Manipulations

Ninety male undergraduate *S*s met in groups of 6 to 10 persons and worked at sorting IBM cards by hand. *S*s had been obtained through volunteering, but

---

[2] An alternative interpretation, formally similar to the derivation of resisting forces presented here, is that coercion results in the reduction of other kinds of power. If personal attraction is another basis for power (French & Raven, 1959), the use of punishment might lead to a decrease in attraction and consequently to a decrease in conformity.

TABLE 1

NUMBER OF *S*s IN EACH EXPERIMENTAL CONDITION

Sequence of Manipulations	50¢ Fine Is Legitimate	25¢ Fine Is Legitimate
Evaluation too slow, threat, evaluation too slow, 50¢ fine	33 (Legitimate Fine)	22 (Illegitimate Fine)
Evaluation too slow, no threat, evaluation too slow, no fine	11 (No Fine)	10 (No Fine)
No evaluation, no threat, no fine	10 (Control)	4 (Control)

when scheduled for their appointments they were told they would receive $1.00 for working on an actual project at the Institute for Social Research.

The coercive power of the supervisor (a collaborator of *E*) was set up by telling *S*s that he could legitimately fine them some part of their pay if they did not perform their job satisfactorily. The amount of the supervisor's *legitimate coercive power* was varied in *E*'s initial statement to *S*s. Some of the groups were told that the supervisor was allowed to fine them up to 50¢, while others were told that the maximum fine was 25¢. Thus, the supervisor's subsequent 50¢ fine was legitimate in the former groups but illegitimate in the latter. The second independent variable in the 2 × 3 design was the supervisor's use of *coercive power*, as shown in Table 1. For all *S*s the intervening variables—*punishment force*, *attraction to the supervisor* and *resisting forces*—were measured by a questionnaire after the completion of the task. Finally, the dependent variables of *conformity* were measured from *S*s' production records and from their own evaluations of their production on questionnaires between work periods.

Experimental Instructions

E's introduction. *E* first presented a cover story intended to give a plausible reason for having asked *S*s to come, yet to draw their attention away from the critical manipulations. *E* said that he was studying "different aspects of production in the performance of routine tasks" including the degree of communication among workers, the length of rest periods, and the nature of time orientation. Accordingly, "in this group" there would be no communication among workers, long rest periods, and a metronome beating during the work session.

Frank (1944) has shown that *E* has a wide domain of legitimate power that prevents resistance. In order to give *S*s the freedom to resist the supervisor's influence attempts without feeling they were sabotaging *E*'s study, *E* made it clear that his interest in their work was separate from the supervisor's. *E* would be interested only in the effects of the above three factors, and *S*s' work for the supervisor was to be "entirely between you and him."

It was necessary that the extent of the supervisor's legitimate coercive power be set up by an external source. For this reason, *E* explained that since the supervisor "is so late today" he would go ahead and tell *S*s about their pay. They were to get $1.00, but since he had complained about some *S*s' poor work the supervisor had demanded the right to fine for unsatisfactory performance. The supervisor's project head and *E* had together decided that although the *S*s were entitled to their pay, the supervisor should have the right to fine not more than 50¢ (or 25¢, depending on the experimental condition) of an *S*'s pay. Any more than that would certainly be unfair.

E then repeated his instructions about no talking, filling out forms during rest periods, etc. At this time the supervisor entered, apologized for being late, and was introduced to the workers. *E* then left.

Supervisor's directions. The supervisor let it be known that he was not concerned with *E*'s experimental conditions of communication, rest, and distraction; he cared only about getting the sorting job done. He explained that IBM cards had to be hand-sorted in order to separate them by number of punches. He then went on to show *S*s how they were to sort the cards, admitting that the task would be tedious, but insisting that the cards should be sorted correctly and not damaged. *S*s were to work *fast* and *carefully*, but not to worry about individual mistakes since there would be some checking of their work later. Without mentioning any amount of possible fine, the supervisor mentioned that if any fine for poor work would be necessary he would warn the individual concerned beforehand.

Work arrangements. *S*s were then seated at a table separated from each other by boothlike partitions, so that they could neither see each other's work nor conveniently speak to each other. During rest periods *S*s remained seated at their places reading magazines. To abide by the rules of no communication, the supervisor passed out all his evaluations, threats, and information about punishments in written notes. Each *S*'s pay was placed in a small container in his work booth so that the supervisor could administer the fines by actually taking away 50¢.

Schedule of Work Sessions

The schedule of work and rest was as follows:

Work Period 1—6 min.

Rest Period 1—5 min., first questionnaire

Work Period 2—6 min. [base period for measuring change in production following threat]

Rest Period 2—5 min.; No Evaluation *S*s received no communication. No Fine *S*s received evaluation form, "You're working *too slowly*. You won't be fined, but *go faster*." Fine *S*s received evaluation form, "You're working *too slowly*. *Go faster*, or you'll be fined 50¢ during next period." [Second questionnaire]

Work Period 3—6 min. [base period for measuring change in production following fine]

Rest Period 3—5 min.; No Evaluation *S*s received no communication; No Fine *S*s received evaluation form, "Not fast enough. Won't be fined, but speed up." Fine *S*s received evaluation form, "Not fast enough. 50¢ fine. Won't be fined again, but speed up." [Third questionnaire]

Work Period 4—6 min. [at beginning of period, supervisor collected fine from pay containers of Fine *S*s]

Rest Period 4—5 min. [at the end of this period the supervisor told all *S*s, "I'm going out to find *E*, so he can give you his questionnaire. I suppose you might as well keep working for a while."] [Fourth questionnaire]

Work Period 5—4 min. [production in supervisor's absence; *E* stops work and distributes questionnaires when he returns]

Success of the Experimental Manipulation

Three assumptions about the experimental conditions must be stated and tested: (*a*) *S*s perceived the threatened fine as a real punishment; (*b*) *S*s perceived that the supervisor's coercive power existed, i.e., that he was capable of carrying out the fine; and (*c*) *S*s in the 50¢ legitimate fine condition perceived the supervisor's evaluation, threat, and/or amount of fine as more legitimate than did the *S*s in the 25¢ condition who were fined 50¢.

The data support the validity of the first assumption. On the post experimental questionnaire, *S*s who were fined indicated that they would less "like to come back and work on this same task . . . under the same . . . conditions" ($p < .001$, by τ_{bis} one-tailed) than *S*s who were evaluated but not fined. The responses to two other questions were in the same direction but not significant:

> Fine *S*s "like[d] working on the task" less, and expected "people generally . . . to avoid [such] tasks" more, than did the nonfined *S*s.

The second assumption is also supported by the questionnaire responses. Only two *S*s reported after the task that the supervisor had no actual power to fine them.

The third assumption received rather weak support from the data. *S*s in the 50¢ legitimate fine condition attributed more legitimacy to the supervisor on the following questions than did *S*s in the 25¢ legitimacy condition:

> Did you feel your supervisor's evaluations of your work were justified? ($p < .04$, by τ_{bis} one-tailed); In terms of how you were doing on your job, do you think that the supervisor was justified in fining you at all? ($p < .06$, by τ_{bis} one-tailed); Did you think the amount of the fine was fair? ($p < .04$, by τ_{bis} one-tailed); *How much* would you have thought would have been a fair amount for the fine? ($p < .07$, by τ_{bis} one-tailed).

Measures of Conformity

Measures were constructed to reflect *S*s' conformity to the supervisor's influence attempts at three levels of observability.

1. *Most observable conformity.* Since *S*s were told that the supervisor would count the total number of cards they sorted during each period, their "total production" was considered their most overt behavior. For a given work period, for example the fourth, the index of conformity to the speed-up induction was

$$\frac{S\text{'s Total Production in Period 4}}{S\text{'s Total Production in Period 3}}$$

henceforth abbreviated $T4/T3$.

2. *Less observable conformity.* Since *S*s were told that the supervisor would not check the amount of *correct* production between periods, and because such checking appeared to be difficult, "correct production" was chosen as a measure of less overt conformity. For a given work period, for example the fourth, this index was

$$\frac{S\text{'s Correct Production in Period 4}}{S\text{'s Correct Production in Period 3}}$$

abbreviated $C4/C3$.

3. *Least observable conformity.* *S*s' private evaluations of their performance were indicated on forms they filled out during each rest period "purely for the benefit of the experimenter" which it was impossible for the supervisor to see during the experiment. Thus, covert conformity was indicated by the degree to which *S*s changed their self-evaluations in the direction of the supervisor's evaluation of their work. From the questionnaire filled out following the supervisor's evaluation, all *S*s were classified as Accepting, Neutral, or Rejecting.

Measures of Resisting Forces

The measure of resisting forces was obtained on the postexperimental questionnaire from *S*s' responses regarding "factors that tended to decrease your production."[3] *S*s were asked to check as many of a set of alternatives as applied. Included in the list were these alternatives which were taken as resistance, since they referred to the influence attempt of the supervisor as the source of the opposing force:

> I thought the supervisor's evaluation was unfair [Evaluation Unfair]
>
> I resented the threat of being fined [Resent Threat]
>
> I resented being fined [Resent Fine]

Fined *S*s' responses to these three items were significantly associated as indicated in Table 5.

Measure of Punishment Forces

This index was based on three questions answered as part of the postexperimental questionnaire:

> A. After he threatened to fine you, how much did you think the amount of the fine would be?
>
> B. If you failed to comply with his recommendations, what did you think were the chances that the supervisor would fine you? [Expressed in probability.]

---

[3] Additional less direct indices of resisting forces were available from other questions, but only one ("Did you ever feel like getting up and walking out . . . ?") was significantly associated with the more direct measures. This single measure is not reported: in general it was not significantly related to the variables predicted.

C. Just after the supervisor evaluated your work for the first time, what did you think were the chances that you would be able to follow his recommendations enough so that you would avoid being fined? [Expressed in probability.]

The strength of the punishment force was computed by using A as the "valence of punishment," B as the "probability of punishment if nonconformity," and (1 − C) as the "probability of punishment if conformity" terms in the formula on p. 93.

HYPOTHESES AND RESULTS[4]

The first three hypotheses deal with the effects of the punishment force and resisting forces on conformity.

Hypothesis 1. The greater the strength of punishment force induced by O's threat the greater will be P's *conformity following the threat*. This hypothesis applies only to conformity during the time period 3/2 when there is still a possibility of future punishment. As shown in the first row of Table 2, there was a significant relationship in the predicted direction between strength of punishment force and conformity following threat. Hypothesis 1 was confirmed for all levels of observability, despite the fact that the questions from which the index of punishment force was computed referred only to total production.

These findings cannot be attributed to confounding of the punishment force measure with the experimental manipulation of legitimacy or with reported resisting forces. The means of the punishment force measure for *S*s in the two legitimacy conditions, and for *S*s with and without each resisting force, were nearly equal.

Furthermore, it can be derived that threatened *S*s (who, we assume, have a greater punishment force) would conform more than *S*s given an influence attempt without threat. However, since coercion also produces resistance (Hypothesis 5b was supported), this prediction could be counteracted by the resistance. Following the first influence attempt there were greater increases in threatened than in nonthreatened *S*s' mean total production ($p < .05$, two-tailed) and mean correct production ($p < .10$, two-tailed); however, there was no significant difference in acceptance of evaluation.

TABLE 2

RELATION BETWEEN STRENGTH OF PUNISHMENT FORCE[a] AND CONFORMITY IN PERIODS FOLLOWING THREAT AND PUNISHMENT

Time Period	Increase in Total Production	Increase in Correct Production	Acceptance of Evaluation
3/2 (following threat)	$r = +.44^b$ $p < .005$	$r = +.38^b$ $p < .01$	$\tau = +.35$ $p < .004$
4/3 (following punishment)	$r = +.09^c$ p *NS*	$r = -.37^c$ p *NS*^{d}	$\tau = +.02$ p *NS*

[a] Strength of punishment force was computed from question-
naire responses, as explained on p. 96. The calculations are based
on $N = 43$. Only those fined *S*s who answered the relevant ques-
tions, and whose punishment force index was nonnegative, were
included.

[b] The correlation between strength of punishment force, and
"differential conformity," i.e., increase in correct production rela-
tive to increase in total production following threat was +.09,
indicating no significant difference in relation to these two
measures of conformity.

[c] The correlation between strength of punishment forces, and
increase in correct production relative to increase in total pro-
duction following punishment was −.35 ($p < .05$, two-tailed).
indicating a significant effect on differential conformity.

[d] In the wrong direction, as far as evidence for the *rejection of Hypothesis 2* is concerned.

Hypothesis 2. The strength of punishment force induced by O's threat is not positively associated with P's *conformity after the threatened punishment has terminated*. After punishment, conformity is no longer instrumental unless there is some expectation of further punishment, so the original punishment force should no longer influence *P*'s conformity. The results in the second row of Table 2 support the prediction made by Hypothesis 2 of no positive association. Again this relation is not confounded by the effects of resisting forces.

Similarly, Hypothesis 2 would predict that *S*s undergoing threat and punishment do not conform more following the fine than *S*s receiving influence attempts only. This comparison of fined and nonfined *S*s shows no significant differences in conformity following the administration of the fine.

Hypothesis 3. The greater the strength of resisting forces following an influence attempt, the less will be the conformity. The results for nonfined and fined *S*s were analyzed separately, since the test of Hypothesis 1 showed that

---

[4] Directional predictions were derived from the theory for all hypotheses, and therefore one-tailed tests of significance were employed. Where several factors led to doubt about the direction of prediction in supplementary analyses, two-tailed tests were used.

TABLE 3

MEAN CONFORMITY FOR *S*s WITH RESISTING FORCE ($\bar{X}_r$) MINUS MEAN CONFORMITY FOR *S*s WITHOUT RESISTING FORCE ($\bar{X}$)

Resisting Force Measure	*T3/T2* (Following Threat)		*T4/T3* (Following Punishment)	
	Increase in Total Production[a]	Increase in Correct Production[b]	Increase in Total Production[c]	Increase in Correct Production[d]
S reports production decreased because:				
Evaluation unfair	$\bar{X}_r - \bar{X} = -.03$ $t = -.87$ p *NS* $[\bar{X}_r - \bar{X} = -.10]$ $[t = -1.52]$	$\bar{X}_r - \bar{X} = +.06$ $t = +1.18$ p *NS*[e] $[\bar{X}_r - \bar{X} = -.11]$ $[t = -1.06]$	$\bar{X}_r - \bar{X} = +.02$ $t = +.45$ p *NS*[e]	$\bar{X}_r - \bar{X} = -.11$ $t = -2.00$ $p < .05$ $[\bar{X}_r - \bar{X} = -.06]$ $[t = -1.07]$
Resented threat	$\bar{X}_r - \bar{X} = -.06$ $t = -1.91$ $p < .05$ $[\bar{X}_r - \bar{X} = -.10]$ $[t = -1.55]$	$\bar{X}_r - \bar{X} = -.08$ $t = -1.50$ $p < .10$ $[\bar{X}_r - \bar{X} = -.15]$ $[t = -1.55]$	...	...
Resented fine	...	...	$\bar{X}_r - \bar{X} = -.03$ $t = -.87$ p *NS*	$\bar{X}_r - \bar{X} = -.12$ $t = -2.34$ $p < .05$ $[\bar{X}_r - \bar{X} = -.15]$ $[t = -3.06]$

[a] Based on 43 scores corrected for +.44 correlation with punishment force. Figures in brackets are based on uncorrected scores for all 55 fined *S*s.

[b] Based on 43 scores corrected for +.38 correlation with punishment force. Figures in brackets are based on uncorrected scores for 55 fined *S*s.

[c] Based on uncorrected scores for 54 fined *S*s (one walked out after the fine).

[d] Based on 43 scores corrected for −.37 correlation with punishment force. Figures in brackets are based on uncorrected scores for 54 fined *S*s (see Footnote c).

[e] Nonpredicted direction.

conformity was related to strength of punishment forces. In the former group no significant relationships were observed between the one available resisting force measure ("evaluation unfair") and conformity. In the Threatened-Fined group the covariance of production with strength of punishment force was removed where correlations were significant by computing the difference between conformity score and conformity predicted from punishment force. The means of (corrected) conformity scores for *S*s with and without each resisting force are compared in Table 3. Among threatened and fined *S*s Hypothesis 3 received fair support (apparently more strongly for correct than for total production), although the picture is unclear because of differing results for corrected and uncorrected scores (both of which are reported in Table 3).[5] An index of differential conformity was used to measure the ratio of observable conformity to less observable conformity. This $\frac{C4/C3}{T4/T3}$ index, based on uncorrected scores, showed that the difference in conformity for *S*s on the basis of the Resent Fine resisting force was significantly greater ($p < .02$, two-tailed) for *C4/C3* (i.e., covert) than for *T4/T3*. The resisting force measures were unrelated to *S*s' acceptance of evaluation.

Since there is some evidence that the legitimacy conditions differed in strength of resisting forces (see Hypothesis 6b), Hypothesis 3 would predict that conformity would be greater in the Legitimate than in the Illegitimate Fine condition. However, there were no significant differences using either corrected or uncorrected production scores or the measure of acceptance of evaluation, probably because the differences in resistance were too small.

The last three hypotheses deal with attraction, coercion, and legitimacy as determinants of resisting forces.

---

[5] Although observed effects could not be attributed to differences in punishment forces, scores were corrected in the hope of decreasing extraneous variance.

Hypothesis 4. The strength of P's *resisting forces against an influence attempt will vary inversely with his attraction to* O. This hypothesis states that resisting forces are related to *P*'s "dislike" of *O*. *P*'s attraction to *O* was measured by *S*s' responses to the question, "If you had a chance to work for your supervisor on some other part-time or summer job, do you think you would work for him?" (Responses to a more direct measure of attraction, namely "liking," were practically invariate.)

Hypothesis 4 received moderate support from the data. For *S*s in the Fine conditions there were inverse associations between attraction and individual resisting force measures for "evaluation unfair," "resent threat," and "resent fine" of $-.30$ ($p < .02$), $-.11$ (p *NS*), and $-.23$ ($p < .04$), respectively as measured by biserial tau. The association between the total number of such resisting forces and the measure of attraction was $\tau = -.18$ ($p < .04$, $N = 55$). Within the No Fine condition the association between the one resisting force measure and attraction was approximately zero ($N = 21$).

It should be observed that the hypothesis does not test the causal dependence of resisting forces on the relationship with the supervisor.

Hypothesis 5a. The strength of attraction toward O *decreases with threat and punishment.* As predicted, it was found that *S*s in the Legitimate Fine condition were less attracted to O (measured in the same way as in Hypothesis 4) than those in the No Fine condition ($\tau_{\text{bis}} = +.39$, $p < .001$, $N = 54$). This result confirms the previous finding by Raven and French (1958b).

Hypothesis 5b. From Hypotheses 5a and 4 we derive: *the strength of resisting forces increases with threat and punishment.* As indicated in Table 4, *S*s in the No fine condition reported resisting forces less often than *S*s in the Legitimate Fine condition. The difference for resisting forces due to evaluation was not significant. Although there was a highly significant difference for resistance due to the fine, the measure of resistance due to the fine was not entirely appropriate in the No Fine condition where the *S*s received only an indirect threat and were never fined. A more appropriate comparison would have been possible only if the control group received a very small fine instead of no fine at all. Thus the findings are consistent with the predictions and the results from Hypotheses 4 and 5a, but the test of the hypothesis is not very satisfactory.

TABLE 4

PROPORTION OF *S*s REPORTING RESISTING FORCES WITHIN EXPERIMENTAL CONDITIONS

Resisting Force Measure	Experimental Condition		
	No Fine ($N = 21$)	Legitimate Fine ($N = 33$)	Illegitimate Fine ($N = 22$)
S reports production decreased because:			
Evaluation unfair	.24	.27	.55
Resent threat	(.19)[a]	.39	.50
Resent fine	(.05)[a]	.52	.64

[a] *S*s in the No Fine condition received a less direct threat than those in the other groups, and were never fined.

Hypothesis 6a. The strength of attraction toward O *increases with increasing legitimacy of an influence attempt.* Especially where coercion is used, the legitimacy of *O*'s acts should lessen the decrease in attraction which threat and punishment presumably cause. Contrary to the prediction, *S*s in the Legitimate Fine condition were less attracted to *O* as compared with *S*s in the Illegitimate Fine condition ($\tau_{\text{bis}} = +.08$, NS) A further test of the hypothesis by relating attraction to *questionnaire* measures of legitimacy also failed to confirm it (see Table 5). This finding fails to support the two previous experiments by Raven and French (1958a; 1958b), though we are here dealing with legitimacy of punishment rather than legitimacy of role occupancy.

Hypothesis 6b. From Hypotheses 6a and 4 we derive: *the strength of resisting forces decreases with increasing legitimacy.* However, since 6a was not confirmed, we can no longer assume that the effect of legitimacy of punishment on resisting forces is mediated by attraction. The last two columns in Table 4 show that the illegitimate treatment produced more resisting forces by all three measures, but only the "evaluation unfair" measure is significantly different in the two experimental conditions ($p < .05$, χ^2 with Yates' correction). Biserial tau between the total number of resisting forces and the two experimental conditions was $+.20$ ($p < .07$).

A further examination of the evidence for

TABLE 5

ASSOCIATIONS BETWEEN MEASURES OF LEGITIMACY, ATTRACTION, AND RESISTING FORCES

(For fined *S*s)

Variables		Illegitimacy					Resisting Forces (*S* Reports Production Decreased Because)		
		1	2	3	4	5	6	7	8
Variables	Measure	Evaluations of Work Unjustified	Unjustified in Fining at All	Amount of Fine Unfair	Fair Amount of Fine (Small)	Attraction to supervisor	Evaluation Unfair	Resent Threat	Resent Fine
	$N =$	55	55	46*	39*	55	55	55	55
2	$\tau =$	+.74	—	—	—	—	—	—	—
	$p <$	.001							
3	$\tau =$	+.65	+.67	—	—	—	—	—	—
	$p <$	.001	.001						
4	$\tau =$	+.51	+.59	+.59	—	—	—	—	—
	$p <$	.001	.001	.001					
5	$\tau =$	−.03	+.01	+.01	+.01	—	—	—	—
	$p <$	*NS*	*NS*	*NS*	*NS*				
6	$\tau_{bis} =$	+.45	+.36	+.54	+.43	−.30	—	—	—
	$p <$	.001	.003	.001	.003	.02			
7	$\tau_{bis} =$	+.08	+.10	+.39	+.29	−.11	+.37	—	—
	$p <$	*NS*	*NS*	.003	.03	*NS*	<.008		
8	$\tau_{bis} =$	+.24	+.26	+.31	+.08	−.23	+.32	+.40	—
	$p <$	.03	.03	.02	*NS*	.04	.03	.004	

* These measures were not obtained from some *S*s early in the experiment; hence the *N*s for Variables 3 and 4 are less than 55.

Hypothesis 6b, utilizing measures of association between the resisting force questions and the *questionnaire* measures of illegitimacy used in the validation of the legitimacy manipulation (see Method section), indicated significant positive associations. As shown in Table 5, 9 of the 12 biserial tau's were significant beyond the .05 level. These highly significant associations should be regarded with caution, however, because the illegitimacy and resisting force questions were similar in content, although the latter questions appeared in a separate portion of the questionnaire in which *S*s were asked to attribute decreases in production to various causes; in addition the distinction between questionnaire measures of legitimacy and resisting forces is supported by their differential association with attraction (see Table 5).

On the basis of the evidence, especially the former tests which utilized the experimental conditions of legitimacy and illegitimacy, Hypothesis 6b receives some support. Comparing the results of Hypotheses 6b and 4, it can be seen that the measure of attraction and the experimental manipulation of legitimacy are similarly related to the resisting force measures, although unrelated to each other.

SUMMARY

In a laboratory experiment 90 *S*s worked on a card sorting task under a supervisor who created various conditions of coercive power. Thirty-three (Legitimate Fine) *S*s were threatened for working too slowly and later fined the amount they had been told was legitimate; 22 (Illegitimate Fine) *S*s were threatened and fined the same amount which, in this condition, was twice the specified legitimate amount of fine; and 21 (No Fine) *S*s were neither threatened nor fined although they were told they were working too slowly.

It was hypothesized that the supervisor's inductions would set up punishment forces and opposing resisting forces both of which would determine *S*'s conformity. Assuming that conformity to coercive power is instrumental to avoiding punishment, the strength of the punishment force should increase with in-

creasing negative valence of the fine and with increasing probability of avoiding it by conforming. The valence and probability of receiving a fine were measured on a questionnaire. The resisting forces should be determined by attraction to the supervisor (i.e., disliking) which in turn is determined by coercion and the legitimacy of coercion.

The findings supported the following hypotheses:

1. The greater the strength of punishment force induced by *O*'s threat, the greater will be *P*'s conformity following the threat.

2. The strength of punishment force induced by *O*'s threat is not positively associated with *P*'s conformity after the threatened punishment has terminated.

3. The greater the strength of resisting forces following an influence attempt, the less will be the conformity.

4. The strength of *P*'s resisting forces against an influence attempt will vary inversely with his attraction to *O*.

5a. The strength of attraction toward *O* decreases with threat and punishment.

6b. The strength of resisting forces decreases with increasing legitimacy.

The following hypothesis was not directly tested, but the data presented were consistent with the prediction as were the findings for Hypotheses 4 and 5a.

5b. The strength of resisting forces increases with threat and punishment.

The findings did *not* support the following hypothesis:

6a. The strength of attraction toward *O* increases with increasing legitimacy of a coercive influence attempt.

For all of these hypotheses except Number 4 the tests included comparisons among experimentally manipulated conditions and hence permitted an unequivocal distinction between the independent and the dependent variables. For Hypothesis 3 only the correlational evidence was significant.

It is concluded that punishment forces and resistance forces are independent determinants of conformity to coercive power. Although it appears that legitimacy affects resisting forces, the present experiment indicates that such effects are not mediated by low attraction to the supervisor.

REFERENCES

FRANK, J. D. Experimental studies of personal pressure and resistance. *J. gen. Psychol.*, 1944, **30**, 23–64.

FRENCH, J. R. P., JR., & RAVEN, B. The bases of power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Inst. Soc. Res., 1959. Pp. 150–167.

RAVEN, B., & FRENCH, J. R. P., JR. Group support, legitimate power, and social influence. *J. Pers.*, 1958, **26**, 400–409. (a)

RAVEN, B., & FRENCH, J. R. P., JR. Legitimate power, coercive power, and observability in social influence. *Sociometry*, 1958, **21**, 83–97. (b)

(Received July 20, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 102-109

RESISTANCE AND CONFORMITY UNDER REWARD AND PUNISHMENT[1]

SHEILA G. ZIPF
University of California, Berkeley

THE PROCESS of social influence has received considerable study in recent years. Findings indicate that conformity varies with such factors as the characteristics of the influencer, the methods of influence used, the nature of the behavior affected, characteristics of the person receiving the influence and, the relationship between the influencer and the recipient (Asch, 1956; Carlson, 1956; Festinger, 1957; Frank, 1944; Hovland, Janis, & Kelley, 1953; Lippitt, Polansky, Redl, & Rosen, 1953). Two aspects of this area seem to have been neglected, however. One of these is the influence process as it affects *actions*. A majority of the studies have been concentrated in influence on opinions or judgments. The other is systematic consideration of the psychological processes that intervene between external conditions and the conformity behavior to produce differential conformity.

Let us view influence on action as a process in which the motivational systems of *Individual P* are changed by the act of another *Person O*. The process of change in motivation intervenes between the influence act of *O* and the conformity behavior of *P*. In terms of individual motivation, one can readily conceive that a single stimulus event may produce conflicting motivations. Thus, the act by which *O* attempts to influence *P* may not only motivate *P* to conform to *O*'s wish, but may also motivate *P* to resist doing what *O* wants. French and Raven (1959) have distinguished several types of influence and proposed that these types should be studied for their differential effects in motivating the person to resist conforming as well as their effects in motivating the person to conform. The terms Conformity Force and Resistance Force will be used for these two conflicting tendencies caused by an influence attempt. French, Morrison, and Levinger (1960) following Lewin (1938) have suggested a formula for the Conformity Force which combines the valence of the goal and the probability of reaching the goal if one does try to conform.

The present study was designed to investigate some of the factors involved in producing the conformity and resistance forces and the resulting changes in action. In order that two different acts would not be involved, a situation was selected in which the conformity force would lead to an increase in a given action and the resistance force to a decrease in the same action. Because the effect of the resistance force was of central interest, it was necessary to produce conditions for which the conformity force could be equated while the resistance force was allowed to vary.

Rewards and punishments, being basic types of influence, were of primary interest. Furthermore, because monetary rewards and punishments permit a reasonable equating of the formula for the strength of the conformity force and were expected to produce different strengths of resistance force, these two types of influence were selected for comparison.

Some of the fairly stable motivational structures of the person should also affect the extent to which an influence attempt on that person produces a resistance force. Since the personality characteristic most likely to be relevant seemed to be the need for independence from control figures, this was included as a variable in the study.

With valence defined as the attractiveness of the goal for the person being influenced and perceived probability as the probability perceived by the person being influenced, the following postulates were framed.[2]

1. The strength of the force to do Action *X*

---

[1] This paper is based on a thesis submitted in partial fulfillment of the requirements for the PhD degree at the University of Michigan. The research was done under contract with the Office of Naval Research at the Research Center for Group Dynamics. Part of this study was reported at the APA meeting in Washington in August 1958. The author wishes to express her appreciation to John R. P. French, Jr. for his valuable assistance in every stage of the study, to John W. Atkinson and Charles Hubbel for many theoretical suggestions, and to Cathryn A. Levison for her constructive review of the manuscript.

[2] A complete presentation of the theory is given in the original report (Zipf, 1958).

is an increasing function of the valence of the Goal G, where Action X leads to Goal G.

2. The strength of a force to do Action X is an increasing function of the perceived probability that Action X leads to Goal G.
3. The strength of a force to do Action X is an increasing function of the perceived probability that one is able to perform Action X in order to reach Goal G.

These general postulates were used to make derivations for both the conformity and the resistance force.

Briefly, the derivations concerning type of influence and personality as they affect the resistance force were as follows:

> The threat and administration of a punishment should tend to make the entire situation negatively valent. There should be forces to leave the field. These forces should operate as resistance forces, opposing the forces to remain in the field and perform Action X. This should not occur in response to the offer and administration of rewards. For those with strong needs to be independent, conforming to the demands of a control figure should itself be somewhat negatively valent. Therefore, there should be a force to not conform. Those who do not have such strong needs for independence should have a less negative valence, or even a positive valence on conforming. Hence, two factors were expected to increase the resistance force and in turn decrease conformity (*a*) punishment and (*b*) the need for independence.

METHOD

Subjects

One hundred and two female *S*s were recruited from undergraduate classes at the University of Michigan. The measure of need for independence was obtained at this time, by a different person who said he was not related to the study but was just pretesting the instrument. He asked everyone to fill out the questionnaire regardless of whether they were signing up for the experiment or not. *S*s were assigned to the laboratory in groups of two to eight so as to keep the distribution of need for independence scores about the same in each condition. Otherwise, assignment was random.

Task

The task the *S*s were given was to sort IBM cards according to the total number of holes they contained. A supervisor, who was said to be connected with a different project for which the cards had to be sorted, was introduced. She explained that she was interested in getting the job done and that the only effect of the experiment was that the *E* was allowed to give out and collect his questionnaires during rest periods and because he wanted conditions of low communication, she had agreed to make any necessary comments to them in the form of written notes. The work was done in booths. There were five work periods of 4 min. each. A more detailed description of the work situation is given in French, Jr., & Morrison, & Levinger (1960).

Manipulation of Sanction

The supervisor explained that she could not make any check on their accuracy during the session, but by measuring the height of the deck of cards they had sorted could tell the speed at which they performed. She indicated that after each work period she would collect their production, evaluate it for speed and give them an evaluation sheet with her recommendations on it. The evaluations were to be made by comparing their work with that of regular employees on the task. Half of the *S*s were told that they might be fined if their work was not up to standards, but that they would be warned of this possibility before the work period in which it would occur. The other half of the *S*s were told the same thing concerning rewards. All *S*s were told that they would start out with a $1.00 payment which was placed in advance in each of the booths. The sequence of events after *S*s had gone to their booths was as follows:

- Work Period 1
- Rest Period—Cards collected; *S*s told that was a practice period
- Work Period 2
- Rest Period—Cards collected
  - Evaluation: Much too slowly
  - Reward induction: Go faster next time and there will be a 50¢ reward.
  - Fine induction: Go faster next time or there will be a 50¢ fine.
  - Self evaluations made
- Work Period 3
- Rest Period—Cards collected; conformity and resistance forces measured
  - Evaluation—Too slowly
  - Reward induction: 50¢ reward for improvement, 50¢ reward next time if you go faster. [reward given]
  - Fine induction: Improvement, but not fast enough, 50¢ fine. Go faster next time or there will be a 50¢ fine. [fine taken]
  - Self evaluations made
- Work Period 4
- Rest Period—Cards collected; conformity and resistance forces measured
  - Evaluation: A little too slowly
  - Reward induction: 50¢ reward for improvement. No reward next time, but go faster. [reward given]
  - Fine induction: Improvement, but not fast enough, 50¢ fine. No fine next time, but go faster. [fine taken]
  - Self evaluations made
- Work Period 5
- Rest Period—Cards collected; resistance force and

expectation of sanction measured
Evaluation: A little too slowly
Self evaluations made
Final questionnaire administered

Concepts and Measures

The conformity force is defined and measured as valence of the goal, times the probability of reaching the goal by attempting to conform.[3]

Explicitly:

The desire to receive the reward
or avoid the fine ×

The perceived probability that one will be fined or fail to obtain the reward if he does not conform − The perceived probability that one cannot conform enough to avoid the fine or obtain the reward

These components of the conformity force were measured by asking the *S*s how much they wanted to avoid the fine or obtain the reward, and what they thought the probabilities were, on fixed choice scales.

The resistance force is conceived as having the same general form as the conformity force. However, no measures were found suitable to that form and the resistance force was measured in a checklist of "factors leading to lowered production." Items that referred to the behavior of the supervisor, such as "I didn't like the way the supervisor acted," were used. In a list of 22 items, there were 7 such items and the number checked was used as the index of the resistance force.

Conformity was measured by the change in speed following an induction to go faster. The index was: total number of cards sorted in the work period following the induction, divided by the total number of cards sorted in the work period prior to the induction. This index is abbreviated $T\ n/n-1$ where n is the number of the work period for which the induction to go faster was given.

A measure of judgmental conformity was also included. This was on the Evaluation Sheet, on which the supervisor evaluated the *S*'s work on a seven point scale from "much too slowly," to "much too fast." All *S*s received the same evaluation for a given work period. On the same page, *S*s were asked to give their own evaluation of their work on the same scale. The difference between the supervisor's check and the *S*s check is the index of disagreement.

The need for independence was measured by 6 items in a questionnaire. Originally 12 items were selected from a test used by Tannenbaum and Allport (1956) and given to 60 college women. The correlation matrix of the items was obtained and 6 items which clustered and best met the definition of the need to be independent of control figures were used. Since the hypotheses concerning the effects of this need are made for those with a strong need, *S*s referred to as High NI are in the top one third of the experimental *S*s on this measure. Those called Low NI are in the lower two thirds. In all other measures, a break of the distribution into high and low is made as close to the median as possible.

Effectiveness of Manipulations

The main assumption concerning the manipulation of sanctions is that the situation is generally less attractive for the Fine *S*s than for the Reward *S*s. This was tested in several ways. On the question, "How do you feel about the task?" Fine *S*s like the task less than Reward *S*s in each work period for which it was asked (Work Periods 2, 3, 4, 5, $P \leq .05, .06, .05, .08$ respectively, χ^2 test). When asked at the end of the experiment if they ever felt like getting up and walking out, Fine *S*s said they did more often than Reward *S*s ($P \leq .01$, χ^2 test). The assumption appears to be valid.

Another assumption is that the rewards and fines of the same amount may be equated. We found that Fine *S*s do not differ in the degree to which they said they wanted to avoid the fine from the degree to which Reward *S*s said they wanted to obtain the reward, at either time that the question was asked ($P > .30$ one tail t test).

HYPOTHESES AND RESULTS

Hypothesis 1.

The conformity force is positively related to increase in speed following an induction to go faster. This hypothesis is clearly confirmed in data from the fourth work period.[4] The correlations are shown in Table 1. Also in that table, the correlations of each of the components of the conformity force with conformity are given. All of the components are related in the appropriate direction; valence, significantly so. The conformity force is better

---

[3] This specific formula was suggested by French et al. 1960. It is similar both in variables used and in the multiplicative form to the formulations of a number of theorists for the prediction of choice and intensity of behavior (Atkinson, 1957; Lewin, Dembo, Festinger, & Sears, 1944; Rotter, 1954; Siegel, 1957; Tolman, 1955). The probabilities as given in the formula differ slightly from the formal statement of the components of the force in the postulates. This is primarily due to the necessity of stating questions so they could be readily understood by the *S*s.

[4] The hypotheses were generally not confirmed by data from the third work period. A previous study (French, 1960) had obtained some of the predicted results under similar conditions in the third work period. However, 6-min. work periods were used at that time. It appears that a certain amount of practice on the task is necessary before these motivational variables are effective. Only the fourth and fifth work period results will be given in this report. There are, also ten *S*s omitted from the fourth work period data. Their conformity force measures indicated that they did not understand the questions. They have been omitted from the data on the fourth work period for clarity of presentation. In no case does the exclusion of these 10 change whether the result of an hypothesis is significant or not although the magnitude of the effects are generally greater than when they are included.

TABLE 1

CORRELATIONS OF THE CONFORMITY FORCE VARIABLES AND THE CONFORMITY FORCE FORMULA WITH INCREASE IN SPEED IN THE FOURTH WORK PERIOD

(T 4/3)

Variable	r	N	p
A. Valence of the goal	.31	92	.001
B. Probability that one will be fined or fail to receive the reward for nonconformity	.16	92	.06
C. Probability that one cannot conform enough to avoid the fine or obtain the reward	−.17	92	.06
Conformity Force Formula A[B − C]	.44	92	.001
Conformity Force Formula *S*s have Low Resistance Force	.36	63	.001
Conformity Force Formula *S*s have High Resistance Force	.41	29	.01

Note.—All p values given are for one-tail tests.

TABLE 2

RELATIONSHIP OF THE RESISTANCE FORCE TO CHANGE IN SPEED OF SORTING IN THE FOURTH WORK PERIOD

(T 4/3)

			Resistance Force		
All *S*s ($N = 92$)			Low	High	
	T 4/3	Low	27	20	
		High	36	9	$\chi^2 = 4.42$ $p = .02$
			Resistance Force		
Low Conformity Force ($N = 46$)			Low	High	
	T 4/3	Low	16	13	
		High	15	4	$\chi^2 = 1.66$ $p = .10$
			Resistance Force		
High Conformity Force ($N = 44$)			Low	High	
	T 4/3	Low	11	7	
		High	21	5	$\chi^2 = 1.20$ $p = .15$

Note.—All p values given are for one-tail tests.

related than any of its three components, and is significantly as predicted for the total group. Partial correlations of each of the components of the conformity force on T 4/3 have been obtained for the Reward *S*s only. They indicate that each variable is significantly ($p < .01$) related to change in speed at that time, in support of the postulates given above.

Since the conformity force and the resistance force are conceptually similar and are predicted to have opposite relationships to conformity, one can be sure that a result obtained with one is not due to the other only if it can be shown that the two are independent. In the fourth work period, the two forces are not significantly related ($p > .50$, χ^2 test). To remove any possibility of confounding, however, when presenting the results of one of the forces, the other is controlled by holding it at two levels; high and low. This is not complete control. However, since the measure of resistance force is extremely skewed, over 50% of the *S*s having a score of zero, other methods cannot be used. As shown at the bottom of Table 1, the first hypothesis is confirmed when this control is used.

Hypothesis 2.

The resistance force is negatively related to increase in speed after an induction to go faster. These data for the fourth work period are presented in Table 2. The relationship is significant for the total N, but not so when the conformity force is held at high and low. This finding may indicate that the relationship in the total N is caused by the conformity force. However, since the conformity and resistance forces are not related, it seems more likely that the reduction in N caused by the control is responsible for the loss of significance.

The assumption involved in measuring the conformity and resistance forces that have opposite relationships to conformity is that the resultant of the two should predict extent of conformity better than either alone. Because the two forces were not measured in the same units, the resistance force could not simply be subtracted from the conformity forces as would be preferred. Furthermore, the resistance force distribution is not suitable for partial regression analysis. But a simple combination of the two forces is possible. Dichotomizing the distributions, *S*s were grouped according to whether the conformity and resistance forces were both high, both low, or opposite to each other. The relationship of such a combination to increase in speed

TABLE 3

RELATIONSHIP OF THE CONFORMITY AND RESISTANCE FORCES TO CHANGE IN SPEED OF SORTING IN THE FOURTH WORK PERIOD

(T 4/3)

Conformity Force Resistance Force		High Low	High High	Low Low	Low High	
T 4/3	Low	11	7	16	13	
	High	21	5	15	4	$\chi^2 = 8.22$ $p < .05$ (two-tail test)

TABLE 4

COMPARISON OF REWARD AND FINE CONDITIONS ON MEAN RESISTANCE FORCE IN THE FOURTH AND FIFTH WORK PERIODS

Fourth Work Period	Fine		Reward		
	N	$\bar{X}$	N	$\bar{X}$	t
All *S*s	40	1.15	52	.25	3.82[a]
Low Conformity Force	22	.91	27	.35	1.83[a]
High Conformity Force	18	1.44	25	.15	3.59[a]
Fifth Work Period					
All *S*s	47[b]	.77	55[b]	.09	4.50[a]
No Sanction Expected	29	.83	36	.01	15.04[a]
Sanction Expected	9	.56	13	.23	2.99[a]

[a] $p \leq .05$, one-tail test when correction for number of degrees of freedom is made because of a significant difference in variances (Dixon & Massey, 1951).

[b] The total N is greater than the sum of N in the two subgroups because expectation of sanction was not ascertained for some *S*s.

is given in Table 3. When the two forces are opposite to each other, the change in speed is as expected. When both forces are high or low, the increase in speed is as likely to be high as low. This finding lends support to the idea of a resultant force combining the conformity and resistance forces to predict extent of conformity and also makes acceptance of Hypothesis 2 more credible.

Hypothesis 3.

Resistance forces are greater in Fine than in Reward. In the fourth work period, this hypothesis is clearly confirmed, as shown in the top half of Table 4.

Hypothesis 4.

Conformity is greater in Reward than in Fine. This hypothesis was not confirmed. There are no significant differences between the Reward and Fine conditions on increase in speed of sorting at this time, neither when the conformity force is held at high and low, nor with the total N. The hypothesis follows logically from the theory only if Hypotheses 1, 2, and 3 are substantiated. That Hypothesis 2 received questionable confirmation may therefore explain this result. Another factor is that only 47% of the Fine *S*s had a score of greater than zero on the resistance force measure and the mean for them alone was only 2.56 out of a possible 7. It is likely that the resistance forces which were produced by the fines tended to have the predicted effects as indicated by Tables 2, 3, and 4, but that it was not produced in sufficient strength in enough of the Fine *S*s to result in significant differences in conformity between the Fine and Reward conditions.

A secondary test of Hypothesis 4 may be made concerning judgmental conformity, using the measure of disagreement with the supervisor's evaluation. It was found that Fine *S*s disagree with the supervisor more than do the Reward *S*s (p = .05, one-tail t test). The *S*s knew that their self-evaluations would not be seen by the supervisor and could not be sanctioned. On the other hand, speed on the task was specifically being sanctioned. Two possible conclusions are suggested: (*a*) Hypothesis 4 is correct only with respect to judgmental conformity and not for action conformity; (*b*) the resistance force generated by this situation does lower conformity of Fine *S*s below that of Reward *S*s, but only when the nonconformity does not put the resistor in danger of being punished.

The fifth work period was carried out to observe the effects of the resistance and conformity forces after sanctions had been removed. It was expected that Fine *S*s would still have resistance motivation and tend to conform less than Reward *S*s even though there would be no more fines or rewards.

The conformity force was not measured for this work period; instead, a measure was taken of whether or not *S*s expected that there would be sanctions at this time. This measure is not related to conformity in this work period, directly, but it does have an indirect effect that will be discussed presently. The same measure of resistance that was used in the previous work period was given again for the

TABLE 5

COMPARISON OF REWARD AND FINE CONDITIONS ON MEAN INCREASE IN SPEED OF SORTING IN THE FIFTH WORK PERIOD

(T 5/4)

	Fine		Reward		
	N	$\bar{x}$	N	$\bar{x}$	t
All Ss	47[a]	1.00	55[a]	1.00	0
No Sanction Expected	29	.99	36	1.04	2.92[b]
Sanction Expected	9	1.01	13	.99	−.37[c]

[a] The total N is greater than the sum of N in the two subgroups because expectation of sanction was not ascertained for some Ss.

[b] $p \leq .01$, one-tail test.

[c] A negative t indicates a difference in the nonpredicted direction.

fifth work period. The measure is not related to conformity at this time. Its validity depends on whether the checklist contains the appropriate sources of resistance. The measure seems to operate as expected with fourth work-period data, but it is likely that the factors producing resistance motivation in the fourth work period were different from those in the fifth work period, since the situation had changed considerably. The items used, apparently, were not very relevant to conformity on the task in the fifth work period.

The resistance force index does seem to indicate a negative attitude at this point, however, because it again turns out to be higher for Fine *S*s than for Reward *S*s. This is shown on the lower half of Table 4 for the total group, with expectation of sanction held constant.

The experimental groups were compared on conformity for this work period. Again, those who expected sanctions were separated from those who did not. This comparison is given in Table 5. When sanctions were not expected, Fine *S*s conform less than the Reward *S*s. When sanctions are expected, the Reward and Fine *S*s do not differ in conformity.

The results of the fifth work period seem to indicate that some amount of resistance force persisted in Fine *S*s even when there were no more fines threatened. The conformity finding lends support to the second conclusion offered for the results of Hypothesis 4 above. In this work period, it was possible to compare two sets of Reward and Fine groups, one where sanctions were still expected and the other where they were not. For those who do not expect sanctions, resistance evidenced by slowing down on the task does not endanger the resistor with punishment or loss of reward. For those who do expect sanctions, slowing down does endanger the resistor in this way. In the fifth work period when the danger is not present the Fine *S*s conform less than the Reward *S*s, however, when the danger is perceived to be present, the conformity of the Reward and Fine *S*s does not differ.

Figure 1 shows that a similar conclusion might have been made by considering simply the number of cards sorted by each of the experimental groups in each work period. Although the Reward *S*s tend to perform at a higher level throughout, there is a significant difference only after sanctions had ceased, in the fifth work period ($P < .04$, one-tail t test).

A number of questions were asked on the questionnaire given at the end of the experiment. The responses are of interest because they indicate the effects of the experimental manipulation on factors other than conformity, particularly those motivations which were not explicitly included in the measures of the conformity and resistance forces. Each effect mentioned was tested on a two-fold contingency table ($p < .05$, χ^2 test). Some of the findings are that the Fine *S*s liked the task, the situation and the supervisor less than did

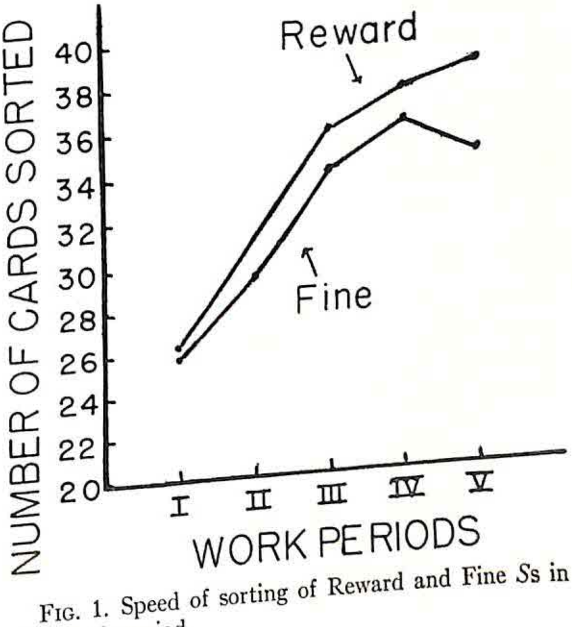

FIG. 1. Speed of sorting of Reward and Fine *S*s in each work period.

the Reward *S*s. They felt the supervisor was more annoying, less competent, and her evaluations less justified. They thought it was less important to the supervisor that they do well on the task, and they felt less accomplishment from going faster. The rewards and punishments thus appear to have affected the satisfactions of a number of motives, such as achievement, aggression, and affiliation, as well as the need for independence, on which more explicit data are available.

It was predicted that, regardless of the experimental condition, those with High NI would have greater resistance forces and less conformity than those with Low NI. This measure does not differentiate individuals on resistance or conformity in the fourth work period. However, in the fifth work period, the difference in the resistance force approaches significance ($p = .07$, one-tail t test, $N = 102$). Comments from *S*s on their questionnaires and in interviews later indicate that the index of resistance force would have been more sensitive to the reactions of *S*s with High NI had it included items referring to the incessant character of the supervisor's demands.

The same comparison was made on conformity of action (T 5/4). Those who are High NI tend to conform less than those who are Low NI but the difference is significant only for those who expect sanctions ($p < .05$, one-tail t test, N = 22). *S*s with High NI who felt that they might be fined or lose a reward if they did not increase their speed on the task are the ones who resisted, and actually slowed down. This finding is rather tenuous, but it suggests an hypothesis for future research; individuals with strong needs for independence as measured by the instrument are aroused to resist conformity by the repetition of sanctioned influence attempts.

DISCUSSION

The purpose of this study was to investigate conformity of actions in terms of individual motivation. Conformity and resistance forces were the basic motivational constructs. These constructs proved useful both for the building of a theory and in the interpretation of results. The conformity force is particularly valuable because of the specificity of its formulation and its predictive ability. It would seem to offer an excellent framework for more basic definitions of different kinds of influence. Types could be defined in terms of the components of the conformity force which are changed and the extent of the change.

The resistance force, though less adequately measured, does have explanatory and predictive value. The weak test made of the effectiveness, as a predictor of conformity, of a resultant force combining the conformity and resistance forces received reasonable support.

More work is needed to improve the measure of the resistance force so that it is composed of the postulated components of a motivational force. This may be done by considering all of the relevant goals in the situation and summing all the forces that are opposite to the conformity force (Carlson, 1956; Thomas & Zander, 1959). An alternative is the way that the conformity force was handled in the present study. In this case a single goal, reward or avoidance of fine, served as a symbolic goal for many nonmonetary goals simplifying the measure. Such improvements in the measure of the resistance force would permit further investigation into the conditions under which varying strengths of the resistance force do or do not produce actions and judgements that are contrary to the demands and judgements of the influencing person.

SUMMARY

An experiment was designed to compare influence by means of rewards and punishments for differential effects on motivation to conform (conformity force) and motivation to resist (resistance force) and the resulting change in actions. Subjects were premeasured on the need to be independent of control figures. The following hypotheses were confirmed:

1. The conformity force taken as a combination of valence of the goal and the probability that trying to conform will lead to the goal, is positively related to conformity of action as measured by increase in speed following a sanctioned induction to go faster.

2. The resistance force is negatively related to conformity of action as measured by increase in speed after a sanctioned induction to go faster.

3. Influence by means of punishment pro-

duces a greater resistance force than influence by means of reward.

The following hypotheses for further research were suggested:

1. When the amount of punishment and reward are equal, persons influenced by punishment conform less than those influenced by reward if the nonconformity behavior is not likely to lead to punishment. This relationship does not hold for nonconformity behavior that is likely to lead to punishment.

2. The motivation to resist produced in those influenced by means of punishment persists for some period of time after punishments have ceased.

3. Persons with strong needs to be independent of control figures are aroused to resist conforming by repeated influence attempts involving sanctions.

REFERENCES

ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority, *Psychol. Monogr.*, 1956, **70**(9, Whole No. 416).

ATKINSON, J. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

CARLSON, E. S. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.

COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512–532.

DIXON, W., & MASSEY, F. *Introduction to statistical analysis*. New York, McGraw-Hill, 1951.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.

FRANK, J. D. Experimental studies in personal pressure and resistance: I. Experimental production of resistance. *J. gen. Psychol.*, 1944, **30**, 23–41.

FRENCH, J. R. P., JR., & Raven, B. The bases of social power. In D. Cartwright, (Ed.). *Studies in social power*. Ann Arbor, Mich.: Inst. Soc. Res., 1959. Pp. 150–167.

FRENCH, J. R. P., JR., MORRISON, H. W., & LEVINGER, G. Coercive power and forces affecting conformity. *J. abnorm. soc. Psychol.*, 1960, **61**, 93–101.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.

LEWIN, K. The conceptual representation and the measurement of psychological forces. *Contr. psychol. Theor.*, 1938, **1**, No. 4.

LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.

LIPPITT, R., POLANSKY, N., REDL, F., & ROSEN, S. The dynamics of power. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. Evanston, Ill.: Row, Peterson, 1953.

MACCORQUODALE, K., & MEEHL, P. Preliminary suggestions as to a formalization of expectancy theory. *Psychol. Rev.* 1953, **60**, 55–63.

ROTTER, J. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

SIEGEL, S. Level of aspiration and decision making, *Psychol. Rev.*, 1957, **64**, 253–262.

TANNENBAUM, A. S., & ALLPORT, F. H. Personality structure and group structure: An interpretive study of their relationship through an event-structure hypothesis. *J. abnorm. soc. Psychol.*, 1956, **53**, 272–280.

THOMAS, E. J., & ZANDER, A. The relationship of goal structure to motivation under extreme conditions. *J. indiv. Psychol.*, 1959, **15**, 121–127.

TOLMAN, E. Principles of performance. *Psychol. Rev.*, 1955, **62**, 312–326.

ZIPF, SHEILA. An experimental study of resistance to influence. Unpublished doctoral dissertation. University of Michigan, 1958.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 110–118

RACE[1], SOCIAL CLASS, RELIGION, AND NATIONALITY AS DETERMINANTS OF SOCIAL DISTANCE[2]

HARRY C. TRIANDIS AND LEIGH MINTURN TRIANDIS
University of Illinois

SINCE Bogardus' (1928) pioneering work with an ordinal scale of social distance, in which a subject indicated zero social distance by stating that he was willing to marry a member of a particular ethnic group, and maximum social distance by stating that he would exclude such a person from the country, social distances scales have been used in a large number of studies. A reliable ranking of ethnic groups was made possible with data obtained with these scales. About 90% of white native-born Americans, it was found, are willing to marry Canadians, or Northern Europeans, but only 10% are willing to marry Southern and Eastern Europeans, and only 1% are willing to marry Negroes. About 10% of these subjects would exclude Jews, Serbo-Croatians, Negroes, Turks, and several other groups from the United States.

After Bogardus' use of the social distance scale with ethnic groups it was used by a number of researchers with other groups. For instance, Wilkinson (1929) used it with occupations, Bogardus (1933) with races, occupations, and religions. Sartain and Bell (1949) published equal appearing interval scale values. Melikian and Prothro (1952) obtained the social distance of Lebanese students towards national and religious groups and found greater social distance towards certain national groups than towards certain religious groups. The reverse result had been obtained by one of these authors in 1935. Prothro and Miles (1953) used the Sartain and Bell (1949) scale in the deep South. Ansari (1956) used the scale in India and found a relationship between social distance and stereotypes similar to that obtained by Katz and Braly (1933) in the United States. Hunt (1956) used the scale in the Philippines and found that his *S*s showed little social distance towards Americans, Christians, and whites, but very much social distance towards Jewish, Moslem, and Buddhist groups. Kirsh (1957) used the scale in connection with voting behavior. Best and Sohner (1956) used it with political, economic, racial, religious, and nationality groups. They found very large social distances towards fascists, anarchists, and communists, large distances towards racial groups, small distances towards religious groups, and very small distances towards national groups. Hill (1953) has criticized the concept of social distance and presented arguments in favor of the usage of interviews and projective techniques for the determination of prejudice, rather than objective indices of social distance, such as the Bogardus scale.

The literature reviewed commonly recognizes that race, religion, nationality, occupation, or political ideology may be determinants of social distance, but does not explore how these factors combine and interact to determine the total social distance score. Furthermore, because of common ethnic stereotypes, these factors, when studied by the usual Bogardus technique, are confounded, e.g., since most Irishmen are Catholic, one cannot tell whether a reported prejudice towards the Irish represents prejudice towards the nationality or the religion; similarly, since Negro stereotypes imply lower class membership, class prejudice may contribute to reported Negro prejudice.

In addition, the literature suggests that some groups may stress *nationality* more than all other factors (e.g., the Lebanese in the 1950s), other groups *religion* (e.g., the Philipinoes), still other groups *race* (e.g., white Americans). Thus, the background characteristics of the subjects (*S*s) may be related to the weight given to the various factors determining social distance.

The present study had three aims: first, to investigate the relative importance of race, religion, nationality, and social class as determinants of social distance for our population of *S*s. Second, to investigate the extent to which the relative importance of race, religion, nationality, and social class, as determinants of social distance, is dependent upon the subject's own background with respect to these factors. Third, to investigate the effect of the *S*'s race, religion, nationality, and social class upon his overall prejudice.

---

[1] Bogardus (1928) used the word *race* to describe the various ethnic groups judged by his subjects. In this paper we intend to use Hooton's well known definition of race as "a subdivision of mankind the members of which are distinguished by possession of similar combinations of anatomical characteristics due to common heredity."

[2] We are indebted to D. E. Dulany, L. G. Humphreys, W. J. McGuire, and I. D. Steiner for a critical reading of an earlier version of the manuscript and for valuable suggestions.

A factorial design, with two levels of race, two of social class, two of religion, and two levels of nationality was used. An analysis of variance of the social distance scores provided the information required by the first aim. Separate analyses of variance of the scores obtained from *S*s of different backgrounds supplied information concerning the relative importance of the various factors for these different groups of *S*s. The total variance of social distance scores, which is a measure of prejudice[3], was used in connection with the third aim. Finally, additional analyses, using protocols of individual *S*s, investigated the effect of individual differences and the extent to which some *S*s tend to use one factor, other *S*s another, and still others various interactions between factors as the basis for their social distance scores.

METHOD

Standardization of the social distance scale. Thirty-six statements previously used by Sartain and Bell (1949) and 24 additional statements were given to 35 Illinois undergraduates who acted as judges. A graphic form of the Thurstone successive intervals procedure was used as described in Edwards (1957, Ch. 5). This procedure permits the development of an equal interval scale. The rank order correlation between the scale values obtained by this procedure and the scale values obtained by Sartain and Bell (using Texas *S*s), for the 36 common statements, was .98. The statements that were finally selected for use are shown in Table 1. The data obtained from the 35 judges and the 15 statements that were finally selected were checked for unidimensionality, and a Guttman coefficient of reproducibility of .90 was obtained.

TABLE 1

STATEMENTS AND THEIR SUCCESSIVE INTERVAL VALUES ON A 100-POINT SCALE

1. I would marry this person.	0.00
2. I would accept this person as an intimate friend.	11.11
3. I would accept this person as a close kin by marriage.	21.50
4. I would accept this person as a roommate or I would date this person.	29.50
5. I would accept this person as a personal chum in my club.	31.10
6. I would accept this person as a neighbor	38.70
7. I would accept this person as my husband's or wife's friend.	40.90
8. I would live in the same apartment house with this person.	49.40
9. I would accept this person as one of my speaking acquaintances.	52.40
10. I would rent property from this person.	57.50
11. I would give asylum to this person, if he were a refugee, but I would not grant him citizenship.	63.10
12. I would not permit this person to live in my neighborhood.	69.70
13. I would not permit this person's attendance of our universities.	81.00
14. I would exclude this person from my country.	95.00
15. I would be willing to participate in the lynching of this person.	97.20

The scale values shown in Table 1 have been arranged so as to cover a 100-point scale with zero corresponding to the *S*'s endorsement of the statement "I would marry this person." This particular statement was not used in the study, however, because we wished to keep the sex of "this person" constant, and specified that it would be male; the statement was thus unsuitable for our male *S*s. All other statements shown on Table 1 were used.

*Subjects. S*s were 86 undergraduates from the introductory psychology class at the University of Illinois.

The questionnaire. An introduction was attached to the questionnaire in which it was stated that it is natural for people to "feel closer" to their relatives and friends than to strangers. *S*s were urged to indicate how close they feel to various people described in the questionnaire. A statement to the effect that these people will often have unusual combinations of characteristics followed, in which certain extreme combinations, such as Portuguese Negro (from Portuguese Angola) were justified on the grounds that such people do in fact exist though the *S* may not be aware of it. *S* was then asked to indicate which religion among the Protestant, Catholic, Conservative or Reform Judaism varieties he considers *most different* from his own personal religious affiliation. He was asked to think of this particular religion when the words "different religion" appear as characteristics of the person described in the questionnaire. The questionnaire was anonymous, but information was obtained about the *S*'s sex, race, religion, ethnic background, and father's and mother's occupation and education.

A Public Opinion Questionnaire consisting of 14 statements[4] from the Adorno, Frenkel-Brunswik,

---

[3] The Total Sum of Squares (TSS) of an individual's social distance scores, towards the 16 stimulus people used in this study is a measure of prejudice, since the larger the TSS the greater the difference in the social distance experienced towards the various stimulus people. This may prove to be a particularly subtle way of measuring prejudice.

[4] The items are taken from Adorno et al. (1950), Pp. 255–257. They are No. 1, 4, 12, 13, 23, 26, 29, 34, 35, 37, 39, 42, 43, and 44.

TABLE 2

MEAN RESPONSES BY GROUP ON A 100-POINT SCALE OF SOCIAL DISTANCE

	Whites ($N = 77$)	Negroes ($N = 9$)
Swedish physician, white, same religion	11.6	23.0
Same religion, Greek physician, white	13.8	20.6
French physician, white, different religion	18.5	15.5
Banker, white, different religion, Italian	20.2	14.2
English coal miner, white, same religion	20.0	23.0
Greek truckdriver, white, same religion	21.0	17.0
Swede, truckdriver, white, different religion	23.2	19.4
White, different religion, Portuguese unskilled worker	32.1	23.0
French physician, Negro, same religion	34.2	11.0
Negro, same religion, Portuguese, civil Engineer	33.3	14.0
Civil engineer, Negro, different religion, French	35.4	14.0
Negro, different religion, Portuguese physician	36.8	19.5
Same religion, French, unskilled work, Negro	43.5	14.6
Negro, same religion, Portuguese, truck-driver	44.4	19.0
Different religion, French, unskilled work, Negro	44.6	19.0
Negro, different religion, Portuguese, coalminer	44.6	19.0

Levinson, & Sanford (1950) F Scale was placed at the end of the questionnaire. Since the present study makes only incidental use of the F Scale and the greatest number of published studies used the original scale rather than the modifications suggested by Sullivan and Adelson (1954), Bass (1956), and Jackson and Messick (1957), it was decided to use the scale in its original form.

The main body of the questionnaire consisted of the description of a person, e.g., "A white, Portuguese, unskilled worker, of the same religion as your own" and 15 statements from Table 1. The descriptions of the person are shown in Table 2. Four additional descriptions were used to assess reliability and are described below. *S* was asked if it was likely or unlikely that he would behave towards the person described (see Table 2) in the way described by the 15 statements of the social distance scale (see Table 1). The characteristics of the person were presented in a random sequence, as can be seen from Table 2. The 15 scale statements were also presented in a different random sequence on each sheet. Finally, the sheets were arranged, within the questionnaire, in different random sequences.

Determination of levels of race, social class, religion, and nationality. For levels of race we used white and Negro. For social class we used data from Welch (1949) and considered physician, banker, and civil engineer to be high, and unskilled worker, coalminer, and truckdriver to be low. For religion, we used the distinction between same religion and different religion. Finally, for nationality we employed the Bogardus (1928) data and considered the English, Swedes, and French high and the Portuguese, Italians, and Greeks as low.

Scoring. Since the unidimensionality of our social distance scale was adequate we scored it as follows: if only "positive statements" (scale value smaller than 58) were checked by *S* as being likely, we took as his score the most extreme positive statement checked as likely. If *S* checked only negative statements (more than 60) we took as his score the value of the most extreme negative statement that he considered as likely. When both positive and negative statements were endorsed, *S* received a score of 58. This method of scoring tended to increase the social distance score of many *S*s and produced a relatively more normal distribution of social distance scores. However, since most of our *S*s did not show much social distance the distribution of the scores was skewed, with a mode on the positive side of the neutral point.

Reliability. Fifty *S*s, taken at random from the 86, were used to compute reliabilities. The responses to the "French physician, white, and of different religion" were found to be correlated .96 to the responses to the "English banker, white, and different religion." The responses to the "Negro coalminer, same religion, Portuguese" and "Negro truckdriver, same religion, Portuguese" were correlated .90. Correlations of .69 and .87 were obtained for "French physician and civil engineer, Negro, same religion," and "White Portuguese and Greek unskilled workers" respectively. The average of these correlations is .855. When Fisher *Z*s are averaged, $r = .88$.

Classification of Ss. All *S*s were college students, thus, potentially middle class. In order to obtain some indications of how social distance is related to *S*s social class we categorized our *S*s as follows: Class I ("upper-middle"): At least one parent had a college education and a highly responsible occupation (e.g., physician, professor, executive in a large corporation). This corresponds very roughly to Warner's upper-middle or lower-upper class. Class III ("lower"): Neither parent had a college education and the father was a manual worker. This class corresponds roughly to Warner's upper-lower class. All other *S*s were classified in Class II ("lower-middle"). The *S*'s classification on ethnic background (high or low) was based on his answers to a question about his ethnic background. If he indicated old American, Northern or Northwestern European backgrounds, he was classified as High, otherwise as Low.

RESULTS

Relative Importance of Characteristics of Stimulus People

The first aim of this study was to investigate the relative importance of race, religion, nationality, and social class as determinants of social distance. Table 3 shows that race is the most important determinant of social distance for white *S*s, while interactions between factors are the important determinants

TABLE 3

ANALYSES OF VARIANCE OF SCORES OF 77 WHITES AND 9 NEGROES

Source	Whites				Negroes		
	df	*MS*	% Variance	*F*	*MS*	% Variance	*F*
Race	1	4.34	77.1	2170***	.10	15.4	—
Social Class	1	.93	16.6	465***	.10	15.4	—
Religion	1	.28	5.0	140***	.00	0.0	—
Nationality	1	.05	0.9	25***	.01	1.5	—
Interactions	11	.002	0.4		.042	67.7	—

*** p less than .001.

for Negro *S*s. In other words, Negro *S*s do not experience social distance towards whites, or persons with lower class occupations, or persons of different religion (e.g., Jews), but if the person in question has a combination of these "undesirable" characteristics, e.g., he is a "white, Jewish, Portuguese, unskilled worker," then Negro *S*s feel some social distance as manifested in a refusal to accept such a person as a close kin by marriage, or as an intimate friend.

For white *S*s race and social class are by far the most important determinants of social distance. They account for almost 95% of the variance of social distance scores. If we return to the previously used examples and consult Table 3, we can clarify the meaning of prejudice towards the Irish and the Negroes. Table 3 suggests that most of the prejudice towards the Irish is religious and not national; most of the prejudice towards Negroes is racial and not social class prejudice.

Background of Subjects and Factors Determining Social Distance

The second aim of this paper was to investigate the extent to which the relative importance of race, religion, nationality, and social class as determinants of social distance is dependent upon the *S*'s own background with respect to these factors. Table 2 shows differences between white and Negro *S*s in mean social distance towards various stimulus persons. White *S*s have a minimum of social distance towards the "Swedish physician of the same religion" and a maximum towards the "Negro, different religion, Portuguese, coalminer." Negro *S*s show maximum social distance towards the "Swedish physician," the "English coalminer," and the "Portuguese unskilled worker," and minimum distance towards the "French physician, Negro, same religion." Table 3 shows a very striking difference between white and Negro *S*s in the importance of the various factors as determinants of social distance. Additional analyses of variance, similar to the analyses presented in Table 3, were computed in which the *S*s were grouped according to their social class, religion, ethnic background, sex, and F Scale scores.[5] These analyses indicate that the relatively less prejudiced groups (see below) emphasized race less than the more prejudiced groups. There was relatively little difference between the groups of *S*s on the weight given to social class, religion, or nationality.

Background of Subjects as Related to Their Prejudice

The third aim of this paper was to investigate the effect of the *S*'s race, religion, nationality, and social class upon his overall prejudice. The Total Sum of Squares (TSS) of an individual's social distance scores, towards the 16 stimulus people, is a measure of prejudice (see Footnote 3). A comparison between the TSS of various groups is shown in Table 4.

Chi square tests computed from other data, obtained from the answers that the *S*s gave on the instruction sheet of the questionnaire, indicate that Catholics consider the Jews rather than the Protestants as most different from themselves ($p < .02$), the Jews consider the Catholics as most different ($p < .01$), and the Protestants generally consider the Jews to be more different ($p < .20$). Almost a third of the Protestants, however, feel closer to the Jews than to the Catholics.

Catholics and conservative Jews have the highest F scores, when compared with the mean F score obtained from the Protestant group ($p < .01$; by chi square). Reform Jews have characteristically lower F scores than the Protestant average ($p < .01$; by chi square test).

---

[5] To save printing costs, these tables have been deposited with the American Documentation Institute. Order Document No. 6317 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

In an attempt to obtain a clearer picture concerning the significance of religion, social class, and national background, we considered the 18 possible subgroups that emerge when these three background factors are considered simultaneously. Tables 5 and 6 present the results of this analysis. The results obtained must be considered to be tentative since we have few cases in each cell. Some cells have only one *S*. The analysis of variance of Table 6 indicates that all three factors account for significant portions of the variance. National background is a particularly important background variable; yet it has not been sufficiently explored in studies of prejudice. The interaction between religion and national background also accounts for a good deal of the variance and nearly reaches the .05 level of significance. Most of the variance is accounted by national background, religion, and social class, in that order. This finding suggests that Westie and Westie (1957) have been working with a relatively less significant variable, social class, while religion, and national background are more important variables. A parsimonious interpretation of these data would, ideally, use a few constructs to account for all the findings, while the interpretations examined by Westie and Westie are attempting to explain only the relationship between social class and prejudice.

TABLE 4

COMPARISONS OF MEAN TOTAL SUMS OF SQUARES OF INDIVIDUALS IN VARIOUS GROUPS

Group	TSS	Group	TSS	Level of Significance (by *F* test)
Whites	5.62	Negroes	0.62	.001
High *F* scorers ($F > 3.5$)	9.32	Low *F*	4.71	.01
Lower class	6.56	Lower middle-class	5.90	.05
Lower class	6.56	Upper middle-class	4.30	.01
Lower middle-class	5.90	Upper middle-class	4.30	.05
Protestants	6.70	Jews	3.48	.01
Catholics	7.85	Jews	3.48	.01
Catholics	7.85	Protestants	6.70	.05
Women	7.16	Men	5.15	.01
High Ethnic background *S*s	7.24	Low Ethnic background *S*s	3.92	.01

Notes.—All groups consist of whites only, except for the first row.

Analysis of Individual Protocols

In addition to the group analyses of variance, we also analyzed individual protocols. The social distance scores of some *S*s appear to be determined mostly by one factor, and those of some other *S*s by another factor. For some *S*s combinations of factors, as evidenced by significant interactions in the analysis of variance of their scores, seem more important than any single factor. These analyses also showed that some *S*s deviate from the general pattern (greatest importance to race, then social class, then religion, and minimal importance to nationality) and show no social distance because of race, but a great deal of social distance because of occupation.

TABLE 5

AVERAGE TOTAL SUM OF SQUARES OF SOCIAL DISTANCE SCORES OF *S*s, CLASSIFIED ON THE BASIS OF THE *S*'S BACKGROUND

Social Class	Catholics	Protestants	Jewish
Upper-middle	High Ethnic = 9.15 ($n = 3$) Low Ethnic = 4.95 ($n = 1$)	High Ethnic = 5.75 ($n = 7$) Low Ethnic = 3.10 ($n = 2$)	High Ethnic = 4.24 ($n = 1$) Low Ethnic = 2.29 ($n = 5$)
Lower-middle	High Ethnic = 14.50 ($n = 4$) Low Ethnic = 7.85 ($n = 3$)	High Ethnic = 7.12 ($n = 10$) Low Ethnic = 3.84 ($n = 1$)	High Ethnic = 5.25 ($n = 7$) Low Ethnic = 2.84 ($n = 7$)
Upper-lower	High Ethnic = 10.00 ($n = 3$) Low Ethnic = 5.40 ($n = 8$)	High Ethnic = 9.70 ($n = 10$) Low Ethnic = 5.25 ($n = 2$)	High Ethnic = 7.90 ($n = 1$) Low Ethnic = 4.26 ($n = 4$)

Note.—High Ethnic = *S*s whose parents come from N or NW Europe; Low Ethnic = *S*s whose parents come from S or E Europe; only white *S*s were used for this table.

TABLE 6

ANALYSIS OF VARIANCE OF DATA OF *S*'S BACKGROUND (Table 5)

Source	*df*	*MS*	*F*	% Variance
Social Class (Sc)	2	14.35	8.7*	16.6
Religion (R)	2	27.20	16.5*	31.4
National Background (N)	1	40.50	24.6**	23.4
Sc × R	4	4.74	2.87	1.1
Sc × N	2	1.40	—	1.6
R × N	2	10.55	6.4	12.2
Error	4	1.65	6.4	3.8
Total	17			

* $p < .05$.
** $p < .01$.

DISCUSSION

For the *S*s of this study race is by far the most important factor determining social distance. Furthermore, social class is a much more significant determinant of social distance than is religion or nationality.[6] Social class is relatively more important for groups that show little social distance and have low F scores than for groups showing a great deal of social distance ($p < .05$). The absolute amount of variance due to social class, as distinguished from the percentage of variance due to social class, is about equal in all our groups. That is, the groups showing greater total social distance do so because they discriminate much more on the basis of race than do other groups.

None of the group data show any interaction effects. On the other hand, the analyses of individual protocol show that several interactions determine significant portions of an individual's social distance. It appears that there is no systematic effect of any particular interaction, over all *S*s, so that the individual interaction effects are eliminated when the group score analyses are computed.

Interpretations of the phenomena of social distance and prejudice have often emphasized psychodynamic factors (e.g., Adorno et al., 1950). Allport (1954) has presented a more adequate analysis employing a multilevel approach in which historical, sociocultural, and situational variables interact with psychodynamic variables and the stimulus itself. On the other hand, Allport does not specify which of these factors is most important. The present data are clearly apposite.

Several of our findings do not require interpretation. That white *S*s show more social distance than Negro *S*s, that high F score *S*s show more distance than low F score *S*s is not surprising. Since all of our stimulus people were male, the greater social distance shown by women may be an artifact of the testing procedure, since it may include the social distance between the sexes.

The remaining differences, however, need some interpretation. Briefly, we find that Catholics are more prejudiced that Protestants, and Protestants more prejudiced than Jews. The middle classes are less prejudiced than the lower class; finally, *S*s of lower ethnic background are less prejudiced than those of higher ethnic background.

Our attempt to interpret these data uses three basic constructs: conformity, cognitive dissonance, and insecurity. Conformity concerns the adoption of the values of the ingroup. This is unquestionably an extremely significant phenomenon; it has been established by a large number of studies, the best known of which is probably Newcomb's (1943). The fact that national background is so important a background variable suggests that conformity may be by far the most important determinant of social distance. Allport (1954) is probably right when he concludes that "about half of all prejudiced attitudes are based only on the need to conform to custom, to let well enough alone, to maintain the cultural pattern" (p. 286).

The second construct, cognitive dissonance, is adopted from Festinger (1957). In applying dissonance theory to prejudice, we take economic competition, with the socioeconomic inequalities that result from it, as the starting point of prejudice. Prejudice is a device that permits justification of the inequalities that result from economic competition. The person who first started competing for something, in the dawn of history, was not prejudiced. However, after establishing an advantage over other human beings, he felt the need to justify it by arguing that his opponents had deserved to lose because of their inherent inferiority. More explicitly, in terms of Festinger's theory, we find on the one hand the existing social and economic inequalities which result in behaviors that Dollard (1957) has described involving economic, sexual and prestige gains; on the other hand there is the American creed, as described by Myrdal (1944). Cognitive dissonance arises when a person behaves in ways that are incompatible with the American Creed ("all men are created equal . . . "), the Christian ethic or other systems of ethics implying equality in the eyes of a supernatural being or the state. Dissonance is decreased when the "exploited" group is regarded as inferior, since the democratic ideal does not then apply to them. With respect to an aspect of anti-Semitism, dissonance may operate in the opposite direction. The Jew is more often in middle-class than in lower-class occupations. This tendency towards upper middle-class occupations may make the Jew appear to be doing "too well", which may lead to the stereotype that he is pushy, shrewd, grasping, intelligent, sly, and aggressive.

---

[6] These findings must be interpreted as referring only to the average range of variables, as they appear in the United States today. If the range of some of the variables is increased, for example, if pagans, atheists etc., were introduced, a different apportionment of the variance would be obtained.

The dissonance theory of prejudice, that we have presented so far, does not account for the fact that within a group of "exploiters", some become bigots, but others become tolerant, and still others militant liberals. It seems that here the psychodynamic factors described by Adorno et al. (1950) become important. The authoritarian personality is characteristic of a person who is threatened, *insecure* (Gough, 1951), highly conservative, rigid and intolerant of ambiguity (Frenkel-Brunswik, 1949). The authoritarian appears to have a need to reduce the ambiguity involved in high dissonance to a much greater extent than the tolerant individual. The insecure person cannot tolerate dissonance, whereas the secure person can tolerate it.

Adorno et al. (1950, p. 385) found that child training practices were related to the development of authoritarian personalities. *S*s scoring high on the F scale reported that their home discipline was harsh, threatening, and arbitrary. Hart (1957) also found that prejudiced *S*s have been subjected to harsh and threatening discipline. Bronfenbrenner's (1958) analysis of the trends in socialization practices in the United States clearly states that

> in matters of discipline working-class parents are consistently likely to employ physical punishment, while middle-class families rely more on reasoning, isolation, appeals to guilt, and other methods involving the threat of loss of love (p. 424).

Also,

> Over the entire 25-year period studied, parent-child relationships in the middle class are consistently reported as more acceptant and equalitarian, while those in the working class are oriented towards maintaining order and obedience" (p. 425).

To sum up the theoretical argument, child training practices may be responsible for the development of a relatively secure, or a relatively insecure personality. The insecure person is usually incapable of tolerating ambiguity. Since in most highly developed societies there are alternative sets of values, which are often contradictory (e.g., equalitarian values may contradict the value that the in-group is better than the out-group), the insecure person attempts to reduce the dissonance generated by such contradictions by fully adopting one of the available ideologies.

With this theoretical scheme in mind we may re-examine the data. First, we find that lower-class *S*s are more prejudiced than middle class subjects. The child training practices used by the working classes are those which are likely to produce insecurity. This is not true for the child training practices of middle-class parents. This finding, then is consistent with the theory that insecure persons are more prejudiced than secure person.

The differences found to be associated with social class are, in turn, consistent with the findings of other investigators. For instance, Westie and Westie (1957) find maximum social distance between lower-class whites and Negroes, and minimal social distance between upper-class whites and Negroes, with middle-class whites falling between lower- and upper-class whites on their social distance scores.

Second, we find greater prejudice in our Catholic than in our Jewish *S*s, with the Protestants in between. Psathas (1957) gave a questionnaire to high school students with Southern Italian and Eastern European Jewish backgrounds, concerning the child training practices of their parents. The one significant difference between the responses of these two ethnic groups was that the Jewish students reported that their parents had a much greater regard for the children's ability to think for themselves. This stress on intellectual independence extended to all areas of thought and Jewish children were taught "to question authority in (their) schooling since even Divine Law is subject to interpretation" (Psathas, 1957, p. 415). This type of socialization is obviously less authoritarian than that of the Catholics. It trains children to tolerate ambiguity, since values are in a state of flux and subject to interpretation. O'Rielly and O'Rielly (1954) found that Catholic *S*s scoring high on a scale of intensity of religious belief were less favorable to Jews and Negroes and favored segregation of Negroes in their own parishes. It would seem, then, that certain features of the socialization practices of Catholic parents, particularly the clarity of their approach towards the Divine Law, predispose their children to intolerance of ambiguity, which in turn is correlated with prejudice. Jews, on the other hand, seem to train their children in a manner which minimizes these tendencies.

The third finding of the present study was that *S*s with Old American, Northern or Northwestern European ethnic backgrounds show more social distance than *S*s with Southern or Eastern European backgrounds. In fact, national background was next to religion our most important background variable. An examination of the analyses of variances of the scores of these two kinds of *S*s reveals that the difference between the two groups with respect to social distance is entirely accounted by the difference in the amount of distance that they experience toward Negroes. This finding cannot be fully interpreted, in the absence of evidence on the child training practices of these groups. Conformity to established cultural norms may be an important factor accounting for it.

The foregoing analysis appears to account for our data, for anti-Negro and anti-Semitic prejudice, with a minimum of theoretical constructs. Like most simple theories it is probably too simple. It seems, however, a reasonably parsimonious first step toward a theory of prejudice. It is limited to an explanation of prejudiced attitudes and does not include an explanation of prejudiced behavior. For such an explanation an extension of the model is required, which in our view should bring considerations from frustration-aggression theory, recently reviewed by Cowen, Lander, & Schaet, (1959), to bear on the probability of occurrence of prejudiced behavior. The theory also provides a series of testable hypotheses about the relationship between prejudice and child training practices, social differentiation, and value conflict.

SUMMARY

In a study of social distance of college students with respect to various social objects, a factorial design with two levels of value of race, social class, religion, and nationality was employed and analyses of variance were computed on social distance scores. For white *S*s race and social class were found to be more important determinants of social distance than religion or nationality. The relative importance of the various factors, however, varied from group to group. White *S*s were found to show more social distance than Negro *S*s; high F score *S*s more than low F score *S*s, women more than men, Catholics more than Protestants, and these more than Jews, lower-class *S*s more than lower middle-class *S*s and these more than upper middle-class *S*s, and *S*s from Northern and Northeastern Europe more than *S*s with ethnic backgrounds from Southern and Eastern Europe. Analysis of the protocols of individual *S*s showed that there is a tendency for the more prejudiced *S*s to assign more of the variance of social distance scores to race than to social class. The data are interpreted in terms of a theory of prejudice that employs conformity, cognitive dissonance, and insecurity as its main constructs.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass: Addison-Wesley, 1954.

ANSARI, ANWAR. A study of the relation between group stereotypes and social distance. *J. educ. Psychol.*, Baroda, 1956, **14**, 28–35.

BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.

BEST, W. H., & SOHNER, C. P. Social distance and politics: Some selected relationships. *Sociol. soc. Res.*, 1956, **40**, 339–342.

BOGARDUS, E. S. *Immigration and race attitudes.* Boston: Heath, 1928.

BOGARDUS, E. S. A social distance scale. *Sociol. soc. Res.*, 1933, **17**, 265–271.

Bronfenbrenner, U. Socialization and social class through time and space. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1958. Pp. 400–425.

Cowen E., Lander, J., & Schaet, D. E. The effects of mild frustration on the expression of prejudice attitudes. *J. abnorm. soc. Psychol.*, 1959, **58**, 33–38.

Dollard, J. *Caste and class in a southern town*. New York: Anchor, 1957.

Edwards, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957.

Festinger, L. *A theory of cognitive dissonance*. Evanston, Ill: Row, Peterson, 1957.

Frenkel-Brunswik, Else. Intolerance of ambiguity as an emotional and perceptual personality variable. In J. S. Bruner & D. Kretch (Eds.), *Perception and personality: A symposium*. Durham, N. C.: Duke Univer. Press. 1949. Pp. 108–144.

Gough, H. G. Studies in social intolerance: A personality scale for antisemitism. *J. soc. Psychol.*, 1951, **33**, 247–255.

Hart, I. Maternal child-rearing practices and authoritarian ideology. *J. abnorm. soc. Psychol.*, 1957, **55**, 232–237.

Hill, M. Some problems of social distance in intergroup relations. In M. Sherif & M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953. Pp. 285–307.

Hunt, C. L. Social distance in the Phillippines. *Sociol. soc. Res.* 1956, **40**, 253–260.

Jackson, D. N., & Messick, S. J. A note on "ethnocentrism" and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132–134.

Katz, D., & Braly, K. W. Verbal stereotypes and racial prejudice. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1958. Pp. 40–46.

Kirsch, A. D. Social distance and voting behavior related to N variables. *Dissertation Abstr.*, 1957, **17**, 2673.

Melikian, L., & Prothro, E. T. Social distance and social change in the Near East. *Sociol. soc. Res.*, 1952, **37**, 3–11.

Myrdal, G. *An American dilemma: The Negro problem and modern democracy*. New York: Harper, 1944.

Newcomb, T. M. *Personality and social change*. New York: Dryden, 1943.

O'Reilly, C. T., & O'Reilly, E. J. Religious beliefs of Catholic college students and their attitudes toward minorities. *J. abnorm. soc. Psychol.*, 1954, **49**, 378–380.

Prothro, E. T., & Miles, O. K. Social distance in the deep South as measured by a revised Bogardus scale. *J. soc. Psychol.*, 1953, **37**, 171–194.

Psathas, G. Ethnicity, social class and adolescent independence from parental control. *Amer. sociol. Rev.*, 1957, **22**, 415–423.

Sartain, A. Q., & Bell, H. V., Jr. An evaluation of the Bogardus scale of social distance by the method of equal appearing intervals. *J. soc. Psychol.*, 1949, **29**, 85–91.

Sullivan, P. L., & Adelson, J. Ethnocentrism and misanthropy. *J. abnorm. soc. Psychol.*, 1954, **47**, 246–250.

Wilkinson, F. Social distance between occupations. *Sociol. soc. Res.* 1929, **13**, 234–244.

Welch, M. K. The ranking of occupations on the basis of social status. *Occupations*, 1949, **27**, 237–241.

Westie, F. R., & Westie, Margaret L. The social distance pyramid: Relationships between caste and class. *Amer. J. Sociol.*, 1957, **63**, 190–196.

(Received July 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 119–123

OCCUPATIONAL CHOICE AND COGNITIVE FUNCTIONING:

CAREER ASPIRATIONS OF HIGHLY INTELLIGENT AND OF HIGHLY CREATIVE ADOLESCENTS[1]

J. W. GETZELS AND P. W. JACKSON
University of Chicago

THE RELATIONSHIP between a person's cognitive style and his orientation to the world is of central importance to the understanding of both cognitive functioning and social behavior. The present study investigates the relationship between cognition, as defined by performance on intellective tasks, and personal orientation, as defined by long-term occupational choice and career aspiration.

The study is part of a research program examining the fantasy productions, school performance, and teacher preferences of two types of adolescents (*a*) those exceptionally high in intelligence but not equally high in creativity, and (*b*) those exceptionally high in creativity but not equally high in intelligence. Results already reported indicated: (*a*) despite striking differences in IQ, the two groups were equally superior in school achievement to the population from which they were drawn; (*b*) despite this equal school achievement, the high IQ students rather than the high Creativity students were preferred by teachers; and (*c*) there were significant differences in the fantasy productions of the two groups, the creative adolescents expressing more stimulus-free, imaginative, humorous, and playful themes.[2] The present investigation examines the relationship between the career choices and the cognitive functioning of the two groups.

The purpose of the investigation is two-fold: At the substantive level, to determine the occupational preferences of the two groups; at the theoretical level, to ascertain the relationship between types of cognition and types of career aspiration. Two guiding empirical questions were posed: (*a*) what is the nature of the long-term occupational goals of the two groups? (*b*) to what extent do perceived "teacher standards" and the criterion "success in adult life" enter as motives in the aspirations of the two groups?

METHOD

The population from which the experimental groups were drawn comprised 449 adolescents (largely from academic and professional families) who made up the total student body of a Midwestern private secondary school. The groups were chosen on the basis of their performance on the following tests:

Standard IQ tests

Either a Binet, a WISC, or a Henmon-Nelson score was available for each adolescent. The scores obtained from the WISC and the Henmon-Nelson were converted by regression equation to comparable Binet IQs.

Five creativity measures

These were taken or adapted from Guilford and Cattell, or constructed especially for the study as follows:

1. *Word Association.* The *S* was required to give as many definitions as possible to fairly common stimulus-words, e.g., BOLT, BARK, etc. His score depended upon the absolute number of different categories into which these definitions could be put.
2. *Uses for Things.* The *S* was required to give as many uses as he could for objects that customarily have a stereotyped function attached to them, e.g., BRICK, PAPER-CLIP, etc. His score depended upon both the number and originality of the uses which he mentioned.
3. *Hidden Shapes.* The *S* was required to find a given geometric form that was hidden in more complex geometric forms or patterns.
4. *Fables.* The *S* was presented with four fables in which the last lines were missing. He was required to compose three different endings for each fable: a moralistic, a humorous, and a sad ending. His score depended upon the number, appropriateness, and originality of the endings.
5. *Make-Up Problems.* The *S* was presented with

---

[1] The research described in this report was supported by a grant from the United States Office of Education. The present report is an expanded version of a paper read at the APA meeting, Cincinnati, September 3, 1959.

[2] J. W. Getzels & P. W. Jackson, "The Highly Creative and the Highly Intelligent Adolescent: An Attempt at Differentiation," Paper read at the APA meetings, Washington, D. C., September, 1958; "The Social Context of Giftedness: A Multidimensional Approach to Definition and Method," Paper read at the American Sociological meetings, Seattle, Washington, August, 1958; "The Meaning of Giftedness: An Examination of an Expanding Concept," *Phi Delta Kappan*, November, 1958, 75–78.

four complex paragraphs, each of which contained several numerical statements, e.g., "the costs involved in building a house." He was required to devise as many mathematical problems as he could that might be solved with the information given. His score depended upon the number, appropriateness, and originality of the problems.[3]

On the basis of the IQ measures and a summated score on the five creativity instruments, the two experimental groups were formed as follows:

1. Students exceptionally high in IQ but not equally high on the creativity measures (hereafter referred to as the High Intelligence Group). These students were in the top 20% in IQ when compared with like-sexed age peers, but *below* the top 20% on the creativity measures. Their mean IQ was 150, with a range from 139 to 179. $N = 28$ (17 boys, 11 girls).

2. Students exceptionally high on the creativity measures but not equally high in IQ (hereafter referred to as the High Creativity Group). These students were in the top 20% on the creativity measures when compared with like-sexed peers, but *below* the top 20% in IQ. Their mean IQ was 127, with a range from 108 to 138. $N = 26$ (15 boys, 11 girls).[4]

Four instruments were used to obtain the data with respect to the two experimental questions:

1. *A Direct Sentence Completion Test.* Among the 60 items comprising this instrument were such career preference items as:

> When I grow up I want to become a(n) . . . .
>
> If the choice of a permanent occupation were left up to me, I would choose . . . .
>
> The work that has the greatest appeal to me is . . .
>
> I always wanted to be a(n) . . . .

The *S* was required to respond "directly" to these items by completing each stub so as to "reflect his own preferences." The responses of the two groups were analyzed for the number and kind of occupations mentioned.

2. *An Indirect Sentence Completion Test.* This instrument was parallel to the Direct test, except that the sentence stubs were written in the third person using boys' names for the male *S*s and girls' names for the female *S*s:

> When Stan (Jean) grows up, he (she) wants to become a(n) . . . .
>
> If the choice of an occupation were left up to him (her), Ken (Kate) would choose . . . . (etc.)

The *S* was required to respond "projectively" to the items by completing the stubs as a "verbal speed test." The responses of the two groups were analyzed for the number and kind of occupations mentioned.

3. *A Parent Questionnaire.* This instrument required the parents to respond to two questions regarding their children's occupational goals as follows:

> Has the pupil expressed interest in a particular career? If yes, specify . . . .

The responses from the parents of the two groups were analyzed for the number and kind of occupations mentioned.

4. *An Outstanding Traits Test.* This instrument contained descriptions of 13 children, each of the 13 exemplifying some desirable personal quality or trait. For example, one child is described as having the highest IQ in the school, another as being the best athlete, another as having the best sense of humor, another as being the best-looking, another as being the most creative person in the school, and so on. The *S*s were required to rank these 13 hypothetical children in three ways: (*a*) on the degree to which they would like to be like them, (*b*) on the degree to which they believe teachers would like them, and (*c*) on the degree to which they believed people with these various qualities would succeed in adult life. The rankings given by the two groups within each category and the relationships between the rankings for the different categories were analyzed.

RESULTS

The findings may be discussed with respect to each of the two questions in turn:

1. Differences between the two groups appear both in the *quantity* and *quality* of occupational goals. The *quantity* of occupational possibilities mentioned is significantly greater for the high Creatives than for the high IQs, the latter group mentioning an average of 1.82 different occupations on the Direct Sentence Completion Test and 3.57 on the Indirect Test, the former group mentioning an average of 2.61 and 5.00 on the respective instruments. The *quality* of the different occupations mentioned as possibilities is also significantly different for the two groups. When the occupations mentioned are analyzed into conventional (lawyer, doctor, professor) and unconventional (adventurer, inventor, writer) categories, it is found that the high Creatives give a significantly greater proportion of unconventional occupations than do the high IQs. For example, on the Direct Sentence Completion Test, 16% of the high IQs mention an unconventional occupation, 62% of the high Creatives mention at least one such occupation as a possibility. The complete data are presented in Table 1.[5]

---

[3] The descriptive data with respect to these tests will appear in the technical report being prepared for the United States Office of Education. In general, the intercorrelations among these instruments for the total population ranged from .079 to .466, with a mean of approximately .30.

[4] The mean IQ for the total student body was 132, with a standard deviation of 15.1.

[5] It is interesting to note that mention of unusual occupations by the high Creatives occurs with equal frequency in their "direct" and their "projective" responses, while for the high IQ's mention of unusual occupations is much more likely to occur in their "projective" responses than in their "direct" responses. This difference supports the proposition held by Kubie

TABLE 1

QUANTITY AND QUALITY OF OCCUPATIONS MENTIONED BY THE EXPERIMENTAL GROUPS ON DIRECT AND INDIRECT SENTENCE COMPLETION TESTS

Test	Group	Number of Occupations Mentioned			Unusual Occupations	
		Total	$\overline{X}$	σ	Number Mentioned	Number of *Ss* Mentioning
Direct	IQ ($N = 28$)	51	1.82*	1.09	6	5**
	Creative ($N = 26$)	68	2.61*	1.41	24	16**
Indirect	IQ ($N = 28$)	100	3.57**	1.81	12	10*
	Creative ($N = 26$)	130	5.00**	1.80	29	17*

t used to test differences between means in Col. 2.
χ^2 used to test differences between frequencies in Col. 5.
* Significant at .05 level.
** Significant at .01 level.

The preceding findings based on the reports of the adolescents themselves are amply supported by the data from the Parent Questionnaire. Parents of 24 of the high IQ children and of 24 of the high Creativity children responded to the Questionnaire. Sixteen of the high IQ parents and 20 of the high Creativity parents reported that their children had expressed an interest in some career. What is most noteworthy in these reports is that 12 or 75% of the high IQ parents mention occupations falling into but *five* career categories: engineering and architecture, science, medicine, law, and teaching; only 7 or 35% of the high Creativity parents mention career choices falling exclusively into these five categories ($\chi^2 = 5.71$, $p < .02$). Instead, the high Creativity parents tend to mention such rather unconventional occupations as "veterinary" or "entertainer," such unresolved combinations as "law or music," "teaching or art," and such expressive professions as "writing" or "dancing," none of these latter being mentioned as possibilities by the parents of the high IQ children.

In short, with respect to the *quantity* of occupations considered, depending on one's point of view, one may say that the high Creativity adolescents as compared with the high IQ adolescents are more "diffuse" in occupational goals or that they are "more able and willing to deal with a greater range of career possibilities." With respect to the *quality* of occupations considered, again depending on one's point of view, one may say that the high Creativity adolescents are more "eccentric" in their occupational goals, or that they are "more able and willing to deal with career risks, that is, to take greater liberties with accepted standards of adult success."

2. Data with respect to the second empirical question, i.e., the "success" orientation of the two groups, were obtained from the Outstanding Traits Test, and provide further insight into the differential career aspirations of the high IQ and high Creativity adolescents.

The high IQs ranked the qualities in which they would like to be outstanding in the following order:[6] (1) character, (2) emotional stability, (3) goal directedness, (4) creativity, (5) wide range of interests, (6) high marks, (7) IQ, (8) sense of humor. The high Creatives ranked the qualities in the following order: (1) emotional stability, (2) sense of humor, (3) character, (4.5) wide range of interests, (4.5) goal-directedness, (6) creativity, (7) high marks, (8) IQ. Most noteworthy here is the extraordinarily high ranking given by the high Creativity group to "sense of humor," a ranking that not only distinguishes them from the high IQ group (who gave it a rank of 8) but also from the student body from which the experimental groups were drawn (who gave it a rank of 7.5).

The most relevant and striking differences, however, between the two groups in the context of the present inquiry are observed in the relationship between the qualities they want for themselves, the qualities they believe lead to adult success, and the qualities they believe teachers favor. For the high IQ group, the

---

and others pointing to the "flux" or movement of preconscious material into the conscious of the creative person, and the relative restriction of such movement in the more highly intellectualized person. Follow-up of these suggestive findings is anticipated.

[6] The entire population of the school almost without exception ranked social skills first as the quality in which they would like to be outstanding, and athletics, good-looks, high energy level, and health last. In view of this very high agreement, the responses were re-ranked omitting the uniformly-ranked qualities. The findings reported here are on the re-ranked data.

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 124–137

SUBLIMINAL PERCEPTION OR PERCEPTION OF PARTIAL CUES[1]

MORTON WIENER AND PETER H. SCHILLER
Clark University

IF THE number of investigations is a valid criterion of importance of an area, "perception without awareness" or "subliminal perception" has been of considerable theoretical as well as of practical concern (e.g., Adams, 1957; Goldiamond, 1958; Jenkin, 1957; McConnell, Cutler, & McNeil, 1958). Interest in this area seems to reflect the influence of personality theories that postulate an ability of the human organism to make discriminating responses to stimuli of which the organism is "unaware" but which are important in determining the behavior of the organism; and the influence of theories that attribute a generally influential role to personality in the determination of perceptual processes. Much experimentation has been initiated with many ingenious approaches, but the findings and conclusions still appear to be equivocal.

Examination and analysis of this vast body of theoretical and empirical data suggests that two basic explanatory models have evolved to account for the phenomena in question (i.e., "perceptual defense," "need perception" and "perception without awareness," etc.). One type of explanation, referred to here for purposes of exposition as the "two-process" view, posits two distinct perceptual processes, a supraliminal one and a subliminal one (e.g., subception). These two processes are integrated, but also operate relatively independently, and different responses may be evoked by each. In this two-process model, a critical stimulus when not perceived via the supraliminal process (consciously) may be perceived via the subliminal process (not consciously); the subliminal process may then set off the appropriate need-related or defense-related responses[2] (Blum, 1955; Klein, 1956; Lazarus & McCleary, 1951; McGinnies, 1949).

The second explanatory model, referred to here as the "one-process" view, posits a single perceptual process that accounts for the phenomena in question as well as for perceptual behavior in general. The one-process view maintains that "perception without awareness" is essentially a misnomer, insofar as total unawareness of the stimulus in subliminal research has not been adequately determined. Rather, the advocates of the one-process view claim that the experiments cited as demonstrating subliminal processes (e.g., Lazarus & McCleary, 1951) from which subliminal processes have been inferred have not adequately investigated the possibility that some supraliminal cues were present. These theorists would hold that it is the perception of these cues that sets off the so-called "need-related" or "defense-related" responses; "need-related" or "defense-related" responses are conceptualized as response hierarchies or response probabilities. Whether one infers "defense" or "sensitization", for example, is a function of a match between the response hierarchy of the

---

[1] The three studies reported here were supported in part by a United States Public Health Research Grant M-2568A to the senior author. Experiment I is based on a thesis completed by Peter H. Schiller under the direction of the senior author and submitted in partial fulfillment of Peter H. Schiller's requirement for an MA degree at Clark University. Portions of Experiment I were presented at the Eastern Psychological Association Meetings, Atlantic City, April 1959. Portions of Experiment II were presented at the Eastern Psychological Meeting, New York, April 1960. The authors wish to thank William Vogel for his help in collecting the data; Eugene Sacks, Kenneth Wagoner, Donald MacLean, Harold Teegarden, Donald Theobald, for help in early pilot work for Experiment II. Experiment III is also part of a series of studies now being carried out by Mr. Schiller designed to try to separate the "difficulty" factor of the intentional tasks from the amount of the intentional task learned and their effects upon incidental learning. We would also like to thank Joachim Wohlwill and Seymour Wapner for their very helpful comments.

[2] The theoretical view that there is "inhibition" of the perception of the noxious stimulus in "perception without awareness" is somewhat more difficult to categorize in the two basic positions proposed here. On one hand, the inhibition model usually posits a single perceptual process; yet on the other, it apparently also posits a "gating" process for channeling the input. In some manner, particular inputs are selectively channeled out, i.e., inhibited from reaching consciousness but are responded to at some other level. In the broadest sense, this would seem to be categorized best with the "two-process" view. (Bruner, 1957; Perloe 1959).

S and the presented stimulus (Goldiamond, 1958; Wiener, 1957).

An elaboration of some of the methodological and logical issues may help clarify the position of each of the two theoretical models. "Perception without awareness" is inferred by those who hold the two-process view whenever the following three conditions occur: (*a*) a stimulus is presented at some intensity (or duration, or clarity) below the level at which *S*s could correctly identify the stimulus on 100% of the trials; (*b*) the *S* cannot or does not correctly identify the stimulus verbally or motorically (consciously); (*c*) some response occurs (e.g., autonomic, or imagery) that would be expected to occur only if the stimulus were somehow perceived and identified.

In psychophysics, "threshold" and "subliminal" are arbitrary statistical concepts. Threshold is usually defined as some intensity of stimulation (or duration) where the *S* can correctly identify the stimulus (or its presence) a certain percentage of the time, usually 50%. Thus, in psychophysical research when the threshold point is identified, it is usually determined as the point where the *S correctly* identifies the stimulus 50% of the time and does *not correctly* identify the stimulus 50% of the time, over a group of trials. Subliminal designates any point of stimulus presentation below this arbitrary point (e.g., from 0% through 49% correct).

Further, because a homogeneous patch of light has usually been the stimulus for establishing the threshold in vision research, it has been acceptable to infer that no portion of the stimulus is perceived if the *S* is unable to identify its presence correctly. However, when a complex stimulus (i.e., a word, or a picture, or a design) is used and presented at a statistically determined threshold point, there is some question whether it is equally reasonable in this instance to assume that *no portion* of the presented stimulus has been perceived, even when the response is incorrect. For when a complex stimulus is presented at any specified threshold intensity, theoretically at least, any one of the following perceptual experiences could occur on *any* single trial: (*a*) the presented stimulus is supraliminally perceived, and the stimulus is correctly identified either by verbal or other motor response; (*b*) some portion of the stimulus is supraliminally perceived, and the stimulus is correctly identified; (*c*) some portion of the stimulus is supraliminally perceived, and the stimulus is incorrectly identified; (*d*) the stimulus is not supraliminally perceived, nor is any portion, and the stimulus is correctly identified (e.g., by chance); (*e*) the stimulus is not supraliminally perceived, nor is any portion, and the stimulus is incorrectly identified. It is important then, in this area of research, to differentiate clearly whether subliminal implies only an intensity of stimulus presentation that results in some percentage of correct report of the stimulus less than 50%, and incorrect identification of the stimulus, or *no* awareness at all by the *S* of the stimulus or of any portion of it.

At issue, then, is whether perception without awareness has been inferred from the occurrence of both *c* and *e* alternatives above, or only on the basis of Alternative *e*. The one-process view assumes that the experimental demonstrations from which "perception without awareness" has been inferred by the two-process view are due to the effects elicited by occurrence of Alternative *c* above; the two-process view assumes that the occurrence of *e* underlies the conclusion drawn with respect to "perception without awareness."

However, despite the possibility of fluctuation of threshold and variations in threshold level for different figures or words, most experimenters have failed to control or assess the threshold level for *each* trial, and therefore the exact perceptual conditions for every trial is uncertain. The one-process theorists maintain that the *S*, on some of the "incorrect" trials at least, perceives some portion of the stimulus, perhaps not sufficient to make a correct report, but that the *S* makes a response which is some portion of the total correct response he would have made if he had perceived all of the stimulus (Ericksen, 1956), and it is these responses to partial cues, that have been attributed to the subliminal process. Some evidence has been cited to support this latter conclusion. It has been demonstrated (Bricker & Chapanis, 1953; Murdock, 1954) that although the first guesses of *S*s are incorrect, under tachistoscopic presentation, correct second guesses are significantly better than could be expected by chance. These investigators conclude that the *S*s must have seen something, although not enough to lead to a correct verbal guess. Voor (1956) found that the subception effect decreased as less information (i.e., lower brightness of stimulus) was available to the *S*.

Thus, to hold the view that subliminal effects occur (two-process model) requires an assertion of the negative proposition, there is *no* evidence of supraliminal effects that can account for the noted behavior. To hold the one-process view requires that evidence be presented that supraliminal cues are present in these instances. It would, therefore, seem the responsibility of those who hold the one-process view to demonstrate that supraliminal effects *are* present and can account for the findings that have been interpreted as evidence of subliminal processes; and the responsibility of those who hold the two-process view to rule out, as carefully as possible, any plausible supraliminal cues, when trying to demonstrate subliminal processes.[3]

In the research to be reported, a series of experiments was undertaken, each a paradigm of the type usually used to demonstrate subliminal perception, to explore whether the effects that apparently occur can be used as a basis for inferring a subliminal process, or whether the effects can be accounted for by the alternative view, i.e., perception of partial cues.

These studies attempt to explore carefully the perceptual conditions in these paradigmatic experiments to determine whether partial cues are perceived, or whether it is reasonable to conclude that since there is no evidence of any supraliminal cue perception, subliminal perceptual processes can be inferred.

Experiment I: Subliminal Stimulation in the Choice Behavior Paradigm

In this paradigm, the *S* is generally presented with a stimulus at some intensity below the point at which 100% correct report occurs; most often this point is set at a very low percentage threshold point. The point typically is determined by a series of trials at the beginning of the experiment. However, since thresholds fluctuate for *S*s with practice, fatigue, motivation; and since any of several perceptual possibilities may occur on *any single trial* at any threshold level (as noted above), it is essential that the exact conditions of awareness of the stimulus be investigated for every trial. Unless this control is established, we cannot conclude that the noted effects of change in choice behavior (or problem solving) under low percentage threshold intensity are attributable to trials on which no portion of the stimulus is supraliminally perceived (rather than the possibility that the effects are attributable to trials on which some portion of the stimulus is supraliminally perceived, although incorrect identification occurs).

The purpose of this experiment is to investigate each of the above possible alternatives by systematically exploring changes in choice behavior when a stimulus is presented at a low percentage threshold intensity. Are the changes in choice behavior a function of those trials in which no supraliminal cues are evident, or can the effects be attributed to trials in which supraliminal cues can be demonstrated?

In this experiment, if the effects are attributable to trials on which all or part of the stimulus is supraliminally perceived, it will not be possible to conclude that *no* effects are ever attributable to trials on which *no* portion of the stimulus is supraliminally perceived—a null hypothesis conclusion. Rather, it would be evidence in support of the one-process view which holds that overall effects which have been in the past cited as evidence for subliminal process effects must be questioned, and that demonstration of subliminal effects will have to be shown to occur when there is *no* apparent evidence of perception of partial cue correlated with the noted effects, for each trial.

Method

Apparatus and procedure. Ten ambiguous line drawings, similar to the Carmichael, Hogan and Walters (1932) series, were presented on a screen by means of a projector. Beneath each design were two words, either of which could be the name for the ambiguous design. The task for the *S* was to choose the word that seemed more applicable. Wiener, Carpenter, & Carpenter (1957) have shown that the choice behavior of a *S* can be modified by this procedure.

In order to control word order choice effects, two sets of slides were prepared. The second set was identical to the first in all respects, except for the reversal of the two words below the figures. On the screen, the figures and the words appeared white on a black background.

---

[3] It is apparent that "subliminal perception" or "subception" cannot be "proven", i.e., it is not possible to prove *no* supraliminal effects are present; nor can the presence of positive evidence of supraliminal processes *disprove* the two-process view either, i.e., there are *no* effects due to a subliminal process.

Two projectors were employed in the study. The first was used to expose the ambiguous designs and the words beneath them. Both designs and words were exposed well above absolute threshold. The second was used to project the subliminal stimulus, a circle, so that it surrounded one or the other word appearing underneath the ambiguous designs. For different experimental groups the circle was projected either with a constant exposure or with intermittent exposure to investigate whether the effects were different for these two types of stimulation. For the constant presentation a conventional 100 w. projector lamp was employed; the power supply to this lamp was put through a transformer. By varying the transformer output and by use of filters, the intensity level of the projector light could be varied. For the intermittent presentation a tachistocopic neon light was placed into the light compartment of the projector. The light was operated by means of a timing unit which was triggered off three times per sec. by a synchronous motor and microswitch.

The screen was 15′3″ from the projector. In the experimental room four desk chairs, each 9′3″ from the screen, were secured to the floor in a semicircle. A 25-w. shielded light was placed 11 ft. above the floor, over the desks. This light stayed on throughout the experiment.

Each ambiguous design was presented for a period of 15 sec., followed by a 15-sec. interval. The *S*s recorded their choices after each presentation, and then checked one of the following three statements for each choice:

(*a*) definitely saw a circle
(*b*) possibly saw a circle but not sure
(*c*) definitely did not see a circle

Before the appearance of each slide a buzzer was sounded to alert the *S*s to the next exposure.

The instructions to the *S*s were as follows:

> In this experiment I am going to show you some slides. Each slide will show a picture of an object. Below the object you will see two words, like you see on this card in my hand (sample card shown). You are to indicate which of the two words is correct for the picture shown on the slide. This experiment was performed before, under slightly different conditions, on a large group of students. They were shown the same pictures, but for a full minute. For each picture that group decided which one of the labels was correct. This time we will show you the same pictures for a somewhat shorter period, and we want to see which label you select. In order to help you to make a choice, sometimes the answers found correct by the other group will be circled, like this [sample card shown]. In some cases the circle will be shown quite clearly. In other cases not so clearly, and still other times the circle will not be shown at all. If you do not see the circle, you will have to make your decision without considering the other group's choice. I will give each of you a sheet of paper now. As you can see, after each number there is a line. After a slide has been shown, write the answer you think is correct on this line. Also, next to the line there are three statements. Please circle the one that applies to you in each case. Please watch the screen during the entire time the picture is shown. Write down your answer and check the appropriate statement only after the picture has been removed. The appearance of each new slide will be indicated by the sound of a buzzer. Do you have any questions? Remember, whenever you see a circle around one of the words, the circle means that that word was chosen by our other group as the correct label. I will turn the light off now. Above you there is a little light which is just bright enough for you to see by. Please go ahead now and fill out the two lines on top of your sheet.

Subjects. 144 undergraduate college students were divided into six groups: two control and four experimental groups. Each of the six groups were further subdivided into two groups: one was shown the ambiguous designs with the choice words in one order; the other subgroup had the order of choice words reversed.

The six groups were as follows:

Control Group I—The circle was never presented.

Control Group II—The circle was presented during each 15 sec. exposure at an intensity where it was plainly visible.

Experimental Group I—The circle was presented constantly during each 15 sec. exposure at a level (52 v. together with a filtering lens) where a group of sophisticated judges had correctly guessed the location of the circle 95% of the time.

Experimental Group II—The circle was presented constantly during each 15 sec. exposure at a brighter intensity (64 v.).

Experimental Group III—The circle was projected intermittently three times per sec., with each duration lasting .07 sec. This level was determined by another group of sophisticated judges (as in Experimental Group I).

Experimental Group IV—The circle was projected intermittently three times per sec. with each duration lasting .09 sec.

For every group, with the exception of Control Group I, the same word was circled for each stimulus design. The word circled for each design was the one less frequently selected by Control Group I. Since this group was not shown a circle at any time, their choices represented the influence of the ambiguous designs alone.

Results

The score for each *S* was the number of the circled words chosen by that *S* as the appropriate label for the designs. Two types of analyses were performed. In the first, the data for the six groups were treated by analysis of variance and by *t* tests. In the second analysis, all of the choices for the *S*s in the experimental group were analyzed according to the subjective reports for each trial. These data were further analyzed by chi square analyses for each design.

Table 1 indicates that there are statistically significant differences among the groups ($p < .01$) in the number of circled choices chosen. Other comparisons show no statistically significant differences.

Table 2 shows the mean number of circled words chosen by each group, the variance, and the percentage of each type of verbal report. As the percentage of the reports of seeing the circle increased, there was an increase in the mean number of circled choices (the circle was always presented below the absolute threshold level for the experimental group *S*s).

Table 3 shows that there are significant differences between the two control groups as well as among those groups differing considerably in the number of subjective reports of the amount of times the circle was seen. No significant differences are found between Control Group 1 (no circle presented) and those experimental groups who had the circle presented at the lowest levels of presentation, either constantly or intermittently.

To assess the influence of the experimental stimulus (the subliminally presented circle) under each of the three subjective reports, all the choices for the *S*s from all of the experimental groups were combined. The number of circled words chosen under each of the three subjective report categories was determined. Table 4 shows the percentage of choices of the circled words for each of the subjective categories and the two control groups. The Not Seen category results in choice behavior very similar to the Control Group I (not shown a circle), while the other two categories are similar to the Control Group II (always shown a circle).[4]

TABLE 1

ANALYSIS OF VARIANCE OF THE NUMBER OF CIRCLED CHOICES SELECTED BY EACH OF THE SIX GROUPS WITH TWO ORDERS OF PRESENTATION

Source	*DF*	*MS*	*F*
Between Groups	5	18.4	5.09*
Between Orders	1	1	—
Interaction	5	3.8	1.05
Error	132	3.6	

* $p < .01$.

TABLE 2

MEAN NUMBER OF CIRCLED WORDS CHOSEN, VARIANCE, AND PERCENTAGE OF SUBJECTIVE REPORTS

Group	Mean	Variance	Definitely Seen (%)	Possibly Seen (%)	Not Seen (%)
Control Group I	3.33	2.68	.4	.4	99.2
Control Group II	5.62	4.66	99.9	.1	0
Experimental Group I	3.62	2.60	2.5	9.2	88.3
Experimental Group II	4.62	5.48	49.1	16.2	34.7
Experimental Group III	3.96	1.93	16.6	23.7	59.7
Experimental Group IV	5.00	4.27	30.8	17.0	52.2

TABLE 3

T TEST ANALYSIS BETWEEN GROUPS OF THE MEAN NUMBER OF CIRCLED WORDS SELECTED BY EACH GROUP

	Exp. I	Exp. II	Exp. III	Exp. IV	Control II
Control I	.617	2.349*	1.147	3.041**	4.171**
Experimental I		1.821*	.615	2.513**	3.642**
Experimental II			1.202	.692	1.821*
Experimental III				1.894*	3.023**
Experimental IV					1.129

* $p < .05$.
** $p < .01$.

TABLE 4

PERCENTAGE OF CIRCLED WORDS CHOSEN BY THE CONTROL GROUPS AND UNDER THE THREE SUBJECTIVE REPORT CONDITIONS

Group	Per cent
Control Group 1	33.3
Not Seen Group	36.5
Possibly Seen Group	53.5
Definitely Seen Group	52.5
Control Group II	55.8

The results of the analyses by chi square show that for 6 of the 10 designs the number of choices for each design is significantly different ($p < .05$) between the two Control Groups. Choices for 4 of the 10 designs are significantly different between the category Not Seen and the combined Definitely Seen and Possibly Seen. There were no significant differences found in the analyses between the Control Group I and the Not Seen category.

---

[4] In this instance, partial information is apparently equally effective in influencing behavior. This finding is not surprising; uncertain information or partial information would be expected to communicate as much information for choice decision—in a two-choice situation—as total information.

In addition the overall magnitude of these latter chi squares is also the smallest. Four of the 10 chi square analyses are significantly different between the Control Group I and the Possibly Seen category.

In summary the analyses of the data yield the following results:

1. Comparison of Experimental Group I, II, III, and IV with Control Group I would appear to demonstrate that subliminal effects have occurred.
2. However, the mean number of circled words chosen increased significantly as exposure brightness increased.
3. Analysis of the data according to subjective reports shows a significantly greater percentage of choices of circled words on those trials on which the *S*s reported seeing or possibly seeing the circle, than on those for which they reported not seeing it. Under the Definitely Seen *and* the Possibly Seen categories the number of choices approximated the choices of Control Group II (always shown the circle). When the *S*s reported not seeing the circle although it had actually been presented, their choices were not different from the Control Group I, i.e., no circle presented at all.
4. There were no significant differences noted between constant and intermittent presentations of the subliminal stimulus.

Discussion

The results of this study support the hypothesis of "perception of partial cues." The experimental stimulus, i.e., the circle presented below absolute threshold, appeared to be effective in influencing choice behavior only when the subliminally presented influencing stimulus was reported Seen or Possibly Seen. In most experiments, since the circle would not be identified in the Possibly Seen conditions, these would have been considered true subliminal trials. However, there was apparently no effect upon choice behavior when the *S*s reported not seeing the stimulus at all. Thus, under the conditions of this experiment, choice behavior was modified by subliminal presentation (i.e., the circle presented below the absolute threshold) only on those trials in which there was partial or complete verbal awareness of the experimental stimulus. These results do not require the inference of subliminal processes.

The question can be raised whether Possibly Seen is equivalent to "perception of partial cues", particularly when a simple stimulus is presented (e.g. a circle). This question requires some additional research. However, it does not seem unreasonable to assume that uncertainty is based in part at least upon some perception of partial cue. Whether or not these are equivalent, the effects noted in other studies that used this type of stimulation procedure and in which *no* perception was inferred because the stimulus was not or could not be identified, must be more carefully assessed to determine whether the noted effects were attributable to the uncertainty (perception of partial cue?) trials rather than to the Not Seen trials.

EXPERIMENT II: SUBCEPTION OR PERCEPTION OF PARTIAL CUE IN THE "AUTONOMIC DISCRIMINATION WITHOUT AWARENESS" PARADIGM

One assumption in the two-process view of need-perception and defense-perception is that discrimination of stimuli (particularly need and defense stimuli) can take place subliminally, and that such discrimination is quite similar in effect to the discriminations that take place in the supraliminal process. Although there are suggestions that the subliminal process may function somewhat differently than the supraliminal, the two-process view has always implied that the subliminal process can and does distinguish stimuli, and evoke responses appropriate to the need-related, or the defense-related stimuli. Further, this discrimination must be based primarily on the "meaning" of the stimulus. Unless stimulus words were subliminally identified and categorized in terms of "need" or "threat", how could the organism respond to STEAK when hungry, or avoid words like WHORE if repressed for sexual stimuli? Without this ability to identify and categorize somehow a correct response could not be expected. The alternative would be to assume there are special structural characteristics of neutral, need, or defense stimuli, a not very tenable assumption. Whatever the process by which subliminal discrimination occurs, the end result is that the discrimination and the responses follow along dimensions that are similar to those which are usually attributed to supraliminal cognitive processes. If, then, the subliminal process evokes responses to stimuli that belong to the same class (i.e., WHORE or PROSTITUTE) and does not evoke this same response to neutral words (i.e., WHERE or CONSTITUTE), nonclass stimuli, then it is possible to investigate the one- and two-process view by a careful analysis of autonomic generalization behavior, in the "subception" paradigm.

Lazarus & McCleary (1951) apparently demonstrated "subception"—a process by which an autonomic discrimination is made to stimuli presented below absolute threshold, even when the *S* is unable to make a correct supraliminal discrimination. Other investigators (Rubenfeld, Lowenfeld, & Guthrie 1956) showed that the subception effect involves stimulus generalization.

Riess (1940) has demonstrated that when verbal stimuli are presented supraliminally the galvanic skin response (GSR) generalizes more to words that are semantically related to the originally conditioned word than to words structurally similar to the supraliminally conditioned words, but there is some generalization to both.

The purpose of this experiment is to investigate the GSR generalization gradients to words semantically similar to the conditioned word (e.g., ship–boat) as compared with words structurally similar (e.g., ship–shop) under conditions of low percentage threshold presentation and incorrect identification of the stimulus.

If it is maintained that the organism "discriminates" a subliminal stimulus and responds to it as belonging to a threat or shock category and the response is greater than to those stimuli which belong to a nonthreat or nonshock category (autonomic discrimination without awareness), then words that mean the same should also show such response. On the other hand, if it is maintained that the effects attributed to subception are attributable to perception of partial cue, (e.g., Bricker & Chapanis, 1953; Voor, 1956) then it is expected that the GSR response should occur to structurally similar words rather than to the words that have the same meaning but are structurally dissimilar. Further, the amount of response to the subliminally presented stimuli should be expected to be a decreasing function as the amount of stimulus perceived by the subject decreases.

Method

Material. The Galvanic Skin Response was continuously measured and recorded with a Grass Model #5 Polygraph. The electrodes were two penny-size sterling silver discs, coated with silver chloride and embedded in foam rubber and strapped to the *S*'s palm.

Shock was administered from 1½ v. dry cells through a Student's Inductorium to electrodes strapped to the leg of the *S*.

The tachistoscope was an electronically triggered, two-field apparatus, built by Roy Gunter Associates of Clark University.

The stimulus words were printed on a tape and presented tachistoscopically.

All of the recordings of GSR and presentations of the stimuli were automatic. A 2 cpm motor and a series of microswitches made it possible to (*a*) present an alerting signal (a red light), (*b*) trigger the tachistoscope at the manually set exposure level, (*c*) mark the time of the stimulus presentation on the recording tape, (*d*) administer shock if scheduled, (*e*) present a verbal response signal (green light), after which the *S* rested until the next alerting signal, and (*f*) move the next stimulus into the tachistoscopic field. This cycle was completed every 30 sec. The shock was administered approximately two sec. following the presentation of the conditioned stimulus.[5] The response signal was presented six sec. after the onset of the conditioned stimulus.

The stimulus words (printed in capitals) were:

Shock	*Semantically Similar*	*Structurally Similar*	*Nonshock*
START	BEGIN	STATE	COULD
SHIP	BOAT	SHOP	EDGE
CENTER	MIDDLE	RECENT	SECOND
ALLOW	PERMIT	FOLLOW	LAUGH
NEARLY	ALMOST	MERELY	DURING

These words were carefully matched for frequency of usage in the Thorndike-Lorge word count. Pilot work indicated there were no apparent differences in thresholds for any category, nor for any particular word or words.

Subjects. The *S*s were all volunteers who were members of a college fraternity. In return for the *S*s' participation, the fraternity was paid for the time they spent in the experiment. In this way, the *S*s felt they were contributing their efforts to research, yet they recognized they were receiving some payment for the discomfort of participating in a shock experiment. Each *S* had the option of having the shock level lowered if he felt it was too painful, and of stopping at any point in the experiment. Three *S*s did not complete all phases of the experiment.

---

[5] Although the literature (White & Schlossberg, 1952; Wickens, 1959) suggests that maximal conditioning occurs if the unconditioned stimuli follow the conditioned stimulus by approximately .5 sec. Pilot work indicated that there appeared to be rapid extinction of the conditioned response under a low level aperiodic reinforcement schedule on those trials on which there were no reinforcement with .5 sec. delay interval.

Twenty-three *S*s completed all portions of the study, but only 12 records met the final three criteria established for testing the hypothesis.

Procedure. To enhance the possibility of conditioning, the *E* was curt, somewhat gruff and impersonal from the time the *S* entered the experimental room until the *S* was finished. It has been discovered in pilot work that the more anxious the *S*, the more easily he seemed to condition to the shock. The *S* was seated in a comfortable chair. Shock apparatus was then attached to his left leg. The GSR recording electrodes were attached to the *S*'s left palm. The *S* was seated behind a cardboard screen. He could see no equipment except the eyepiece of the tachistoscope. A clip board hung near the eyepiece.

Conditioning phase. A card with 10 words, 5 of them with asterisks, was hung on the clip board. The *S* was given the following instructions:

> If you look into the T scope you will see a green glow. When that green glow changes to a red glow, it means, pay attention, for a word will flash on the screen you see back here in the tachistoscope. You are to wait until the red glow changes to green, and then you are to tell me what the word was. In other words, when the red glow flashes, pay attention; when the green glow comes back on, report what the word was and relax. No word will appear in the T scope that does not appear on this card. When a word marked on this card with an asterisk appears, you may or may not be shocked; at first you will be shocked every time a word with an asterisk on this list appears, after that, you may or may not be shocked when a word with an asterisk appears. You will never under any conditions be shocked when a word appears which is not marked on the card with an asterisk. Understand? Let's go.

Conditioning was on a 100% reinforcement schedule for two presentations of each shock word; thereafter, it was an aperiodic 50% schedule. After four presentations of all of the five shock and the five nonshock words, a test trial was given. Just prior to this test trial, the *S* was asked if he could recall the five shock words. A shock was administered as soon as the *S* recalled all five shock words. The criterion of conditioning was that each shock word elicited a higher GSR than each nonshock word on the test trial. If this criterion was not met, the procedure was continued with intermittent test trials following further conditioning trials. In this conditioning phase, words were presented at a supraliminal level, although an attempt was made to keep the exposure time relatively short. The conditioning period lasted between 40 min. and 2 hr.

Subliminal test phase. After a rest period of 5 to 15 min., the *S* was given the following instructions:

> Now I am going to present the words at very fast speeds. You are to tell me what the word was when the green light comes back on. You may not see the word every time; nevertheless *guess*. That's important. If you're not sure, *guess*. Guess from this card [a new card was tacked up on the clip board] There are 20 words here, 10 of them are new; the other 10 you've already seen. You will be shocked, if you will be shocked, only on those words to which you were shocked before. By the way, you haven't been reacting too well; I'm going to raise the shock.

The shock was not in fact raised unless the *S* had shown by his GSR reactions that he had adapted to the shock intensity. Each of the 20 words was presented five times at a point as close to the *S*'s 50% level of recognition as possible. The approximate exposure level was determined by early pilot work. This level was adjusted for each *S* individually as the experiment progressed. Whenever the *S* had a run of trials on which he correctly identified the stimulus, the stimulus exposure duration was decreased; when the *S* had a run of incorrect identifications, the duration was increased. These changes were necessary to keep close to the 40–60% threshold point. The exposure level for all classes of words was not different, since the words from each class were randomly ordered in the list. An aperiodic shock schedule was administered through this phase. There were two shock trials for every five exposures of the shock words.

Supraliminal test phase. During this phase *S* was simply told, "Now I'll show you words at slower speeds, so you can recognize each one." Although no shock was given, the *S* was not so informed and each *S* appeared to expect to be shocked. Each of the 20 words was shown once at a clearly supraliminal level.

Results

Verbal responses. For the 23 *S*s who completed all phases of the experiment, it was possible to examine the incorrect verbal guesses for the subliminal test phase trials. For each shock word (conditioned stimulus), there were two generalization words, one meaning, one structure. Incorrect verbal guesses could be analyzed for the mean number of times a shock word was incorrectly guessed as being its matched meaning word (e.g., SHIP presented—guessed BOAT) .30; or as its matched structure word (e.g., SHIP presented—guessed SHOP) 1.57; and for the number of times the meaning words were incorrectly guessed as its shock pair (e.g., PERMIT presented—guessed ALLOW) .39; and the number of times the structure words were incorrectly guessed as its shock pair (e.g., FOLLOW presented—guessed ALLOW) 2.70. Chance is .30. These differences are reliable. Twenty-one *S*s have more matched pair shock-structure and structure-shock guesses compared to their meaning-shock and shock-meaning guesses. The other two *S*s had equal numbers of incorrectly guessed pairs. Each of these two *S*s had two pairs in the two categories of incorrect guessing.

Autonomic responses. Three criteria were established to determine whether or not each *S*'s record would be analyzed to test the hypothesis. First, over all trials the stimulus words should have been presented between the 35–65% threshold level. One *S* who correctly identified 71% of the stimulus words was excluded as not meeting this criterion. The second criterion was that in the supraliminal test phase, the GSR for at least four of the shock words had to be greater than the GSR to *any* of the nonshock words. Seven of the 23 *S*s did not meet this criterion. The third criterion was that the mean GSR evoked by the shock words had to be greater than the mean GSR evoked by the nonshock words for those trials of the low percentage threshold presentation when the words were correctly identified by the *S*. Three *S*s were excluded as not meeting this criterion. Earlier investigators suggested that these three criteria were minimal requirements to test the hypothesis of this study. Voor (1956), for example, had noted that the conditioning seemed to disappear when the transition was made from relatively slow exposure speeds of the conditioning trial to the relatively fast exposure speeds of the stimulus presentations in the low percentage threshold level test phase. It was evident that if the generalization hypothesis were to be tested, it would be necessary to insure that conditioning was present for each *S* throughout the experiment. The latter two criteria imply that at least minimal conditioning was present for each *S* included in this analysis. Twelve *S*s remained who met all three criteria.

A Percentage Change Score was computed for each stimulus presentation. Three measures of resistance were determined, one at 1 sec., one at 2 sec. and one at 3 sec. prior to the presentation of the stimulus. Each of these readings was converted to conductance units $\left(\frac{1}{R}\right)$. The base level score then for each trial was the average of these three readings. The Percentage Change Score was determined by taking the maximum deflection reading within the 7 sec. following the stimulus presentation (usually between the 4th and 5th sec.), converting this reading to conductance units, and subtracting the computed base level score from this value and then dividing this difference by the base level and multiplying by 10^4 to remove decimals.[6] If the maximum deflection after stimulus presentation was less than the average base level score, the Percentage Change Score was estimated as zero for that trial.

TABLE 5

ANALYSIS OF VARIANCE OF THE PERCENTAGE CHANGE SCORE FOR THE THREE PRESENTATION CONDITIONS AND FOR THE FOUR CLASSES OF STIMULI

Source	*DF*	*MS*	*F*
Between Conditions (C)	2	11,512	3.06
Between Stimuli (S)	3	56,811	3.31*
Between Individuals (I)	11	184,566	9.48**
C × S	6	19,461	9.64***
I × C	22	17,178	8.51***
I × S	33	3,761	1.86*
I × C × S	66	2,019	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Each *S* was then given an average Percentage Change Score for each of the four classes of stimulus words (i.e., shock, meaning, structure, neutral) for his Supraliminal Trials, for those low percentage threshold presentation trials on which he correctly identified the stimulus word, and for those low duration presentation trials on which he did *not* correctly identify the stimulus word. Thus each *S* had 12 scores. These scores were analyzed by an Analysis of Variance. Because the interactions were significant, the first order effects were tested against the interactions. Table 5 indicates that all effects were significant, except Between Conditions, (i.e., Supraliminal, Correct Subliminal, and Incorrect Subliminal trials). However, the Conditions by Stimulus interaction was statistically significant ($p <$ .001). Examination of Fig. 1 shows that the three other classes of stimuli elicited greater responses than did the neutral words, and that

---

[6] Percentage Change Score-

$$\frac{\text{Maximum conductance score} - \text{Base level conductance score}}{\text{Base level conductance score}} \times 10^4$$

the response to the meaning words is not less than to the structure words on the Supraliminal trials. The effect changes on the correct identification trials of the Subliminal presentation. On the trials where the *S*s incorrectly identify the words under low duration presentations, the response to the structure words is greater than all of the other three categories. Noteworthy, too, is the successive decrease in the amount of response to the shock words as the amount of identification decreased (and inferentially the amount of stimulus the *S*s saw of the word).

It has been noted that the GSR is a function of the verbal response, as well as the response to the stimulus. To be certain that these effects cannot be attributed to different guessing responses to structure words than to meaning words in those trials in which stimuli presented subliminally were incorrectly identified, a further analysis was performed on the scores for each *S* for each stimulus category (i.e., shock, meaning, structure, neutral words) categorized by whether the *S* had guessed it as a shock or as a nonshock word. Because six of the 12 *S*s did not have entries in the cell for shock word presented, incorrect shock word guessed, this category and that for shock word presented, nonshock word guessed incorrectly were excluded from the analysis. Three *S*s failed to have a single cell entry in one of the six remaining cells for each *S*. These values were estimated by successive approximations, taking into account both row and column values. The three estimated values were entered in the analysis of variance, and the degrees of freedom for the analysis were reduced by three. Table 6 shows the results of this analysis of variance. Since none of the interactions was significantly different from zero, the sums of squares and degrees of freedom were combined into a pooled error term. Significant *F* values were found between stimuli ($p < .05$) and between individuals ($p < .001$). The variances attributed to kind of guesses (shock word or nonshock word) did not quite reach a significant level (*F* value 3.06; 3.29 needed for $p = .05$). Examination of the means of the mean GSR to the stimulus words indicated that the noted significant differences among the means can be attributed to the higher GSR evoked by the structure words than by the other two categories. Further examination showed that the amount of the *S*s' responses to the structure words was greater for both kinds of guesses—incorrect guessed shock and the incorrect guessed nonshock—although this difference is much greater in the former instance. Figure 2 shows these relationships. It appears clear that the effects attributed to subliminal processes appear to be a function of the GSR responses to the incorrectly identified structure word presentations.

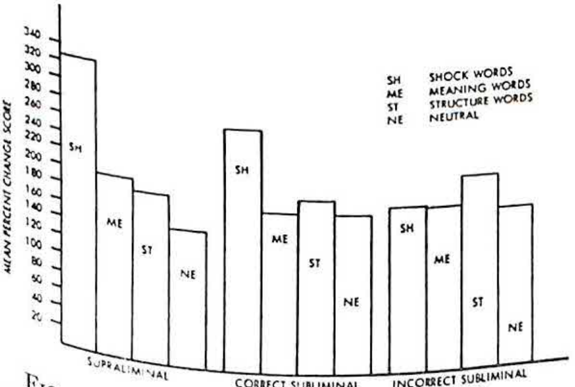

FIG. 1. Mean per cent change score of GSR for each stimulus category for the three stimulus conditions.

TABLE 6

ANALYSIS OF VARIANCE OF THE PERCENTAGE CHANGE SCORE OF THE THREE STIMULUS CLASS WORDS GUESSED SHOCK AND NONSHOCK FOR THE INCORRECT SUBLIMINAL TRIALS

Source	*DF*	*MS*	*F*
Between Guesses (G)	1	6,347	1.37
Between Stimulus Words (S)	2	18,686	4.03*
Between Individuals (I)	11	55,943	12.05**
G × S	2	13,935	2.46
G × I	11	2,218	—
S × I	22	4,128	—
Residual	19	5,662	
Pooled Error	54	4,642	

* $p < .05$.
** $p < .001$.

Discussion

These results also support the hypothesis of perception of partial cues. Under the conditions established to assess more carefully the presence of supraliminal cues, effects that might have been attributed to subliminal processes can be attributed to perception of supraliminal partial cues. In addition not only is there the generalization of the GSR to

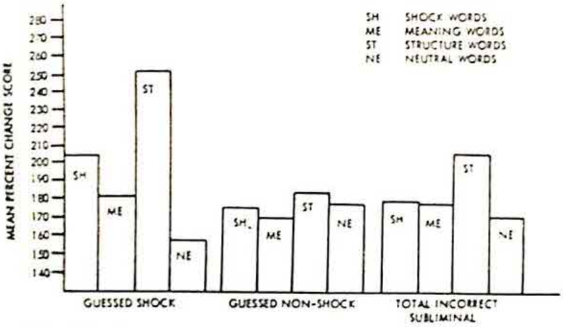

FIG. 2. Mean per cent change score of GSR for each stimulus category analyzed by type of responses for the incorrect subliminal trials.

structurally similar words (Fig. 2), but also the amount of response (Fig. 1) to the shock word decreased as the degree of subjective clarity of stimulus presentation decreased (i.e., from Supraliminal trials, to Correct Identification trials of Subliminal Presentation to Incorrect Identification trials of Subliminal Presentation). Further, the incorrect verbal "guessing" behavior indicated that the *S*s appeared to be responding verbally as well as autonomically to perception of portions of the stimuli.

The findings in this experiment bear upon another issue—whether or not subception in this paradigm is an experimental artifact (e.g., Ericksen 1956). The argument which suggests that subception effect may be an experimental artifact can be described as follows:

> Let us assume that conscious discrimination of the stimulus [verbal behavior] and the autonomic response evoked by the stimulus are correlated, but not completely, i.e., they do not correlate perfectly, even if both response systems were permitted to respond the same way, either as a continuous response or as a discrete response. [The issue is the same under conditions where one response is measured in discrete units, the other in continuous units.] This assumption requires only that it is possible for either or both responses to be right or wrong [by any criterion] or for one or the other to be right on any one trial. If this one assumption is tenable, then only four behavioral possibilities can result:
>
> (*a*) Verbal discrimination is correct—autonomic discrimination is correct
> (*b*) Verbal discrimination is correct—autonomic discrimination is incorrect
> (*c*) Verbal discrimination is incorrect—autonomic discrimination is incorrect
> (*d*) Verbal discrimination is incorrect—autonomic discrimination is correct

In research investigating the operations of subliminal processes, possibilities (*a*) and (*b*) are immediately excluded from further analysis because the organism is "aware" of (identifies) the stimulus and thus these trials are not considered as true subliminal presentation trials. Only possibilities (*c*) and (*d*) would be considered subliminal trials. Possibility (*c*) is not important for determining differences between the two response categories, because it contributes equal error variance to both systems—verbal and autonomic. Thus, since only (*d*) trials are contributing differences, given enough trials for (*d*) to occur, greater autonomic discrimination without verbal awareness could always be found. Because the conditions of analysis have been fixed, whether the stimuli are presented supraliminally or subliminally, autonomic responses will always seem to be more accurate than the verbal responses, and it will always be possible to conclude that subliminal processes perceive stimuli of which the organism is unaware, if the only trials analyzed include only trials on which the verbal (aware) response are incorrect. Whatever the merits of this analysis, the findings in this experiment suggest that the autonomic responses that occur and were interpreted as a subception effect need not be artifacts. Rather, they appear to be a generalization response to the perception of some part of the original conditioned stimulus. If they were artifacts, there would be no reason to expect the autonomic response to correlate more to one of the classes of stimuli (structure words) than to any of the other classes of stimuli (i.e., the other nonshock words, neutral or meaning) under the conditions of (*d*) trials.

EXPERIMENT III: SUBCEPTION OR PERCEPTION OF PARTIAL CUE IN THE PERCEPTION OF PERIPHERAL STIMULI PARADIGM

Some theorists who hold the two-process view maintain that the *S* "perceives" peripheral stimuli of which the *S* is apparently unaware at the time of exposure (e.g., Klein, 1956) but which have later effects upon the *S*'s behavior. Thus, for example, the concept of "registration" could be included in this discussion. That is, the stimulus affects the organism and is "received" by the *S*, but for some reason, does not reach the level of awareness, or consciousness, but still has some subsequent effect—the locus often unspecified. Klein (1956) describes the concept as follows:

> First the experiments show that *registration* must be distinguished from the different *experiential qualities* in which we may be aware of things and events. The processes responsible for phenomenal imprinting seem to be *independent* (this italics ours) of attention-giving processes—the means by which we become aware of things in a distinctive quality. Thus, registration and perception would seem to involve distinct processes (p. 11).

It would seem that registration for some theorists at least is considered a process different from, although an integral part of, perceptual processes in general.

On the other hand, if it is held, as a one-process theorist does, that effects attributed to phenomena such as registration (a paradigm of effects of unaware perception of peripheral stimuli) can be attributed to perception of partial cue, and are only quantitatively different from perception with awareness, it is necessary to demonstrate that effects attributed to "registration" can be shown to be related to the "amount" of partial cues perceived. This formulation of the problem essentially involves an incidental learning paradigm. Previous research has shown that the amount of incidental learning is affected by such variables as the type of orienting and peripheral tasks (e.g., Postman & Adams, 1956); the incentive of the *S* in performing the intentional task (e.g., Bahrick, 1954); and the conditions of presentation of the intentional material (e.g., Neimark & Salzman 1953). In this experiment the incidental learning (assumed to be a measure of the effects of perception) will be investigated under conditions of increasing difficulty of the intentional (i.e., orienting) task.

The third experiment, then, tests the hypothesis that the greater the amount of the peripheral stimulus *S* can perceive, the more registration occurs, even though the stimuli are objectively presented and perceptually equally available to the *S*s. In question in this hypothesis is the view that registration effects found are dependent upon supraliminal perception, as suggested by some theorists. Again, evidence that the effect of peripheral stimulation is a function of the amount of such stimuli perceived does not negate the possibility of registration as a phenomenon. Rather, it calls for closer consideration before the negative proposition is asserted, there is no evidence of supraliminal perception of the peripheral stimulation in registration. In this paradigm, the amount of peripheral stimuli available to the *S*s was varied by controlling the difficulty of an ongoing task, while the peripheral stimuli were exposed. The assumption was made that the more difficult a task (although not so demanding as to exclude any perception of peripheral cues), the less possible it would be to perceive peripheral stimuli.

Method

Subjects. Thirty undergraduate *S*s from an Introductory Psychology class were randomly assigned to one of three groups of 10 *S*s each.

Material. The test stimuli were nine irregular geometric forms. Two sets of these nine figures, randomly ordered (no design was repeated within each set of nine designs) were used. These 18 designs were pasted on a large circular board which rotated with the tracking boards. As the board rotated at a constant speed, each design appeared for a fixed time period in a window opening slightly above the place where the *S* was trying to track on the brass wire tracking pattern. Each of the three tracking boards had a different tracking pattern; one relatively straight and easy, the second more complex, and the third most complex and most difficult. The tracking board made eight complete constant speed revolutions in the 2 min. Each test figure appeared in the window opening 16 times for each *S*, and for the same amount of time for every *S*, no matter which task he was required to trace. The test sheet had the nine geometric designs which ap- appeared on the rotor board interspersed randomly among 33 other irregular geometric patterns, all dittoed on a single sheet of 8½" × 11" paper.

Procedure. The *S* was seated in front of the pursuit rotor and the following instructions were read to him:

> I would like you to take this stylus. Behind these (pointing to the window covering) there is a window. In the window, you will see a brass line. When I start the apparatus, the brass line will start moving, and your task will be to keep the tip of the stylus on the brass line. The length of time you will be able to keep contact will be timed. [Pause] [Opening window cover] Do you see the brass line? O.K. Make contact with it now. Are you ready? Go.

After the *S* completed the pursuit rotor task, he was shown a sheet of paper on which were drawn 42 geometric figures, 9 of which were the same as the 9 exposed to the *S*s on the pursuit rotor task. The *S* was given the following instructions:

> While you were trying to keep the stylus on the brass line, did you notice that there were some figures going by? I'm going to give you a sheet of paper on which you will see a large number of figures. I would like you to pick out those nine figures which you think appeared up there.

TABLE 7

ANALYSIS OF VARIANCE OF THE NUMBER OF PERIPHERAL FIGURES CORRECTLY IDENTIFIED BY *S*s IN THE THREE TASK GROUPS

Source	*DF*	*MS*	*F*
Between Conditions	2	19.00	6.418*
Within	27	2.96	

* $p < .01$.

Results

The score for each *S* was the number of correct designs selected from the 42 possible choices. An Analysis of Variance was computed for these scores for the three groups (i.e., Difficult, Moderate, Easy). Table 7 shows a statistically significant F value ($p < .01$) between the mean scores for the *S*s in the various task groups. The mean correct identification for the Difficult Tracing Group was 1.8; Moderate 2.9; Easy 4.1.

Discussion

Again these results are consistent with the expectations derived from a one-process view and do not appear to be consistent with the expectations suggested by the two-process view that registration is a separate process. If the assumption is tenable that registration is independent of attention to the stimulus, as implied by some theorists, then the expectation that the amount of registration should be constant under equal stimulation seems to be a logical consequence. On the contrary, the results seem to show that tasks that demand greater attention and consequently less opportunity to perceive the peripheral stimuli results in less registration despite constant physical exposure, at least as measured by memory and identification of the peripheral stimuli. However as Klein has suggested, the effects may be more peripheral than the ability to identifiy the stimuli at a later time. Only additional research will determine whether these results will be found when other behavioral manifestations are explored.[7]

Conclusions

The findings of the three paradigm experiments are all consistent with a one-process view,—the hypothesis of "perception of partial cues." It is clearly recognized that these three experiments tap but a small range of the effects that have been attributed to subliminal processes. As Klein, Spence, & Holt, (1958) have noted, the effects of subliminal stimulation may be quite subtle and are not measured by the procedures used in these experiments. This latter issue is not in question. If subliminal effects occur they must be validly demonstrated by those who hold this view. What is relevant is that great care must be taken in accepting or asserting the negative proposition that no supraliminal effects are evident, involved in the inference of subliminal effects, without very carefully controlling and assessing the possibility that supraliminal effects *are* present. Greater care and subtlety must be used in ascertaining the presence or absence of supraliminal cues. Our results do *not* lead us to propose the view that there is *no* behavior without awareness, nor do we propose the hypothesis that only stimuli of which we are aware effect behavior. Rather, it is proposed that more careful and systematic exploration of behavior under varying levels of awareness be considered.

It is unfortunate that usage of the term "subliminal" and "awareness" has resulted in much confusion. The problem of subliminal has been discussed above. Consideration of the term "aware" shows equal difficulties and need for clarification. For example, one can be unaware of a stimulus (e.g., proprioceptive stimulation or X-ray); or be unaware of a response (e.g., a particular speech pattern); or unaware of the relationship between a stimulus and a response (e.g., conforming to norms of a normative group—aware of the behavior of the group, aware that we are making the same response, but unaware that conformity pressures have influenced our behavior). (Wiener, Carpenter, & Carpenter (1957); Wiener, 1959) Careful specification of these possible alternatives seems necessary. In addition, there is a tendency among some psychologists to postulate special and independent processes to account for behavior under low percentage threshold presentation conditions—and make these special instances of "unawareness," while finding no need to postulate special processes to account for the many other instances in which behavior without awareness has also apparently been demonstrated (e.g., Giddon & Ericksen 1959; Greenspoon 1955).

---

[7] A further series of experiments are being prepared to test the effects of "registration" on less cognitive type behaviors, e.g., imagery, dreams, etc.

It would seem evident from these considerations and from the results of these experiments that investigations of "subliminal" and "unawareness" have not always carefully examined nor defined what was implied by these terms. To assume that a new or different process underlies variations in behavior under what appear to be variations in conditions, or to infer a new process based upon acceptance of a negative proposition may lead to a proliferation of constructs which will still further confuse and confound our limited understanding of behavior.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

BAHRICK, H. P. Incidental learning under two incentive conditions. *J. exp. Psychol.*, 1954, **47**, 170–172.

BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **56**, 24–29.

BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

BRUNER, J. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152.

CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.

ERICKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80.

GIDDON, N. S., & ERICKSEN, C. W. Generalization of response biases acquired with and without verbal awareness. *J. Pers.*, 1959, **27**, 104–115.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

JENKIN, N. Affective processes in perception. *Psychol. Bull.*, 1957, **54**, 100–127.

KLEIN, G. S. Consciousness in psychoanalytic theory: Some implications for current research in perception. Mimeo. paper read at Freud Centenary Symposium, Brandeis Univer., Massachusetts, 1956.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUZANNE. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–260.

LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness; A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

MCCONNELL, J. V., CUTLER, R. L., & MCNEIL, E. S. Subliminal stimulation: An overview. *Amer. Psychologist*, 1958, **13**, 229–242.

MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.

MURDOCK, B. B., JR. Perceptual defense and threshold measurements. *J. Pers.*, 1954, **22**, 565–571.

NEIMARK, EDITH, & SALZMAN, I. J. Intentional and incidental learning with different rates of stimulus presentation. *Amer. J. Psychol.*, 1953, **66**, 618–621.

PERLOE, S. I. Inhibition as a Perceptual Process. Mimeo. paper read at APA, Cincinnati, 1959.

POSTMAN, L. J., & ADAMS, PAULINE A. Studies in incidental learning: IV. The interaction of orienting tasks and stimulus materials. *J. exp. Psychol.*, 1956, **51**, 329–333.

RIESS, B. F., Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, **26**, 238–240.

RUBENFELD, S., LOWENFELD, J., & GUTHRIE, M. Stimulus generalization subception. *J. gen. Psychol.*, 1956, **54**, 117–182.

VOOR, J. H. Subliminal perception and subception. *J. Psychol.*, 1956, **41**, 437–458.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

WICKENS, D. D. Conditioning to complex stimuli. *Amer. Psychologist*, 1959, **14**, 180–188.

WIENER, M. Perceptual thresholds: Conditions and parameters. Research proposal to United States Public Health Service, National Institute of Mental Health, 1957.

WIENER, M. Some correlates of conformity responses. *J. soc. Psychol.*, 1959, **49**, 215–221.

WIENER, M., CARPENTER, JANETH T., & CARPENTER, B. Some determinants of conformity behavior. *J. soc. Psychol.*, 1957, **45**, 289–297.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 138–140

CRITIQUE AND NOTES

TRUST, TRUSTWORTHINESS, AND THE F SCALE

MORTON DEUTSCH[1]

Bell Telephone Laboratories, Murray Hill, New Jersey

This study is part of a series of experimental studies (Deutsch, 1957, 1958) that have investigated some of the determinants of trusting behavior. In this brief paper, some striking findings relating trusting behavior to scores on the F scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, p. 260) are reported.

Our experimental work has utilized a two-person non-zero-sum game (see Fig. 1) in which the gains or losses incurred by each person are a function of the choices made between two alternatives by one's partner as well as the choices made by oneself. Person I has to choose between Rows X and Y; Person II has to choose between Columns A and B. The amount of (imaginary) money each person wins or loses is determined by the box they get into as a result of their respective choices. For example, if Person I chooses Row X and Person II chooses Column A, they get into the AX box and they each win $9.

If you examine the possibilities of choice for Person I, you will notice that he can win most and lose least by choosing Y. Similarly, Person II can win most and lose least by choosing B. However, if Person I chooses Y and II chooses B, they both lose $9. Both can win only if they end up in the AX box. If Person I is reasonably sure that Person II is going to choose A, he can win more by choosing Y. Analogously, if Person II is confident that Person I is going to choose X, he can win more by choosing B rather than A.

The essential psychological feature of the game is that there is no possibility for "rational" individual behavior in it unless the conditions for mutual trust exist. If each player chooses to obtain either maximum gain or minimum loss for himself, each will lose. But it makes no sense to choose the other alternative, which could result in maximum loss, unless one can trust the other player.

In the study reported here, *S*s were drawn from an introductory psychology course at a local university. Several weeks prior to the experiment, they had filled out the F scale. During the experiment, the instructions to the *S*s about the game were such that they fully understood the implications of any combination of choices that they and the other person might make and they knew that the other person had a similar knowledge of the game. Unlike some of our other experiments (in which the *S*s were induced to assume a cooperative, individualistic, or competitive orientation), the *S*s were given no motivational orientation. They were allowed to assume whatever orientation they wished to assume vis-à-vis the other person, about whom they were given no information. The *S* did not know the identity of the other person and knew that the other person did not know his identity (except that each *S* knew that the participants in the experiment were all students in the same psychology course).

*S*s played the game twice, each time in a different "position" and each time, presumably, with a different person. In the First Position, *S* made his choice first and his choice was presumably announced to the other person before the other made his choice. In fact, the other person was "fictional" and, hence, the *S* was not informed what the "other person" chose after the *S* made his choice. In the First Position the *S* was faced with the decision of trusting the other person or not. In the Second Position, the *S* chose second after he knew the choice of the other person. Here, too, the other person was "fictitional" and the actual *S* was always informed that the "other person" had chosen Row X (i.e., had trusted). Hence, in the Second Position, the *S* was faced with the decision of being trustworthy or not.

RESULTS

Table 1 presents the data concerning the relationship between choices in the two positions. It is clear that the *S*s who were "Trusting" when they chose first tended to be Trustworthy when they chose last; on the other hand, the *S*s who were "Suspicious" when they chose first tended to be Untrustworthy when they chose last.

Before the *S*s made their choices in the first position they were asked to indicate what they expected the other person to choose. Of the 24 *S*s who chose Row X in the first position and Column A in the second position, 21 indicated that they expected the other person in the second position to choose Column A; of the 22 persons who chose Row Y in

---

[1] This research was conducted while the author was on the faculty of New York University under a contract with the Office of Naval Research: Contract NONR 285(10). The United States Government is given permission to reprint this article in full or in part. I am grateful to Leonard Solomon for his assistance in collecting and analyzing the data and to Irving Taylor for making his students available for the experiment.

	A	B
X	(+9, +9)	(−10, +10)
Y	(+10, −10)	(−9, −9)

Person I chooses between Rows X and Y, Person II between Columns A and B. Person I's pay-offs are the first numbers in the parentheses, Person II's are the second numbers.

FIG. 1. Two-person non-zero-sum game.

the first position and Column B in the second position, 17 indicated that they expected the others in the second position to choose Column B. Thus, predominantly a choice of Row X was a "trusting" rather than "masochistic" choice and a choice of Row Y was correlated with an expectation of "untrustworthiness" rather than of "masochism" from the other person.

When the *S*s were in the second position they were asked to indicate what they thought the other persons (who were then in the first position) expected them to choose. The results indicate that 95% of the *S*s who were Trusting and Trustworthy in their own choices and who had also expected trustworthy choices from the other (i.e., a choice of Column A) thought the other person would be trusting (i.e., would expect a choice of Column A from the *S*); only 35% of the *S* who were Suspicious and Untrustworthy in their own choices and who had also expected untrustworthy choices from the others (i.e., a choice of Column B) thought the other person would be suspicious (i.e., would expect a choice of Column B from the *S*). This finding suggests that the *S*s who were Trusting and Trustworthy expected Trustworthiness and Trusting from the others while the *S*s who were Suspicious and Untrustworthy predominantly expected an exploitative orientation from the others (i.e., they expected the other persons to be trusting but not trustworthy). Hence, it is not surprising that, in the Second Position, these latter *S*s would respond to a Trusting choice from another by "taking advantage" of the other.

The data presented in Table 2 indicate a significant relationship between the *S*'s game behavior and his score on the F scale.[2] Almost all of the *S*s with relatively low scores made Trusting and Trustworthy choices, almost all of the *S*s with relatively high[3] scores made Suspicious and Untrustworthy choices, while the *S*s with medium scores tended to choose one way or the other equally often. It is interesting to note that 7 of the 9 *S*s who were inconsistent in their choices in the two positions (i.e., they were Trusting but Untrustworthy or Suspicious but Trustworthy) were in the medium range of the F scale scores for their class.

TABLE 1

RELATIONSHIP BETWEEN CHOICES IN THE FIRST AND SECOND POSITIONS

First Position Choice	Second Position Choice	
	Trustworthy (A)	Untrustworthy (B)
Trusting (X)	24	5
Suspicious (Y)	4	22

TABLE 2

RELATIONSHIP BETWEEN BEHAVIOR IN THE GAME SITUATION AND F SCALE

Game Behavior	F Scale Score			
	Low (1.2–2.2)	Medium (2.3–3.3)	High (3.4–4.4)	*N*
Trusting and Trustworthy	12	10	2	24
Suspicious and Untrustworthy	0	13	9	22
Suspicious but Trustworthy	0	4	0	4
Trusting but Untrustworthy	2	3	0	5

DISCUSSION

The data of Table 1 provide a striking demonstration of the symmetry of the *S*'s behavior in his two complementing roles vis-à-vis the other person. *His behavior toward the other is congruent with what he expects from the other, and also, what he expects from the other is congruent with his behavior toward the other.*[4] The *S*s tend to be trusting and trustworthy or suspicious and untrustworthy in this essentially ambiguous situation with unknown others. This result suggests that the personality predispositions tapped by the experimental game are not simply one-sided internalized orientations toward another or internalized expectations from another but are instead internalizations of a *reciprocal* pattern of interrelationships with another.

---

[2] The point biserial correlation between the scores on the F scale and the two most frequent categories of game behavior indicated in Table 2 is .50 ($p < .001$). The mean of the F scores for the Trusting and Trustworthy *S*s was 2.38; the mean for the Suspicious and Untrustworthy *S*s was 3.37.

[3] It is evident that the "highs" are high only in comparison with the other *S*s in their class; the high scores in this group are similar to the mean scores reported in other groups (Adorno, Frenkel-Brunswik, Levinson, & Sanford, p. 266). When using the F scale as a personality measure, we believe it makes most sense to consider the individual's score in relationship to the scores of others within his cultural milieu.

[4] It is apparent that the *S*s were not behaving in accordance with the ethical injunction of "do unto others as you would have others do unto you", but were rather guided by the dictum of congnitive consistency "do unto others as you expect others to do unto you and expect others to do unto you as you do unto them."

In other words, what appears to be internalized is a *system* of interrelations between oneself and the other, including the norms which prescribe *both* what to expect from the other and how to act toward the other. A similar conception of the nature of personality predispositions is advanced more fully by such authors as Mead (1934) and Parsons (1955).

The F scale data indicate that the game behavior was related to personality predispositions and that it was not determined by "accidental" orientations assumed during the course of the experiment, except in the special sense that vulnerability to "accidental" factors was itself influenced by the personality predispositions measured by the F scale.

The mass of research on the F scale, surveyed in Christie and Cook (1956), indicates that there are consistent differences between Low and High scorers in their behaviors in social situations and in their responses to various questionnaires. The High as compared with the Low scorers tend to be more authoritarian, less intellectually sophisticated, less liberal in their political views, more cynical concerning human nature, more prejudiced toward minority groups, and to have experienced and to favor stricter child rearing practices. The present results suggest that in addition, one may say that in an ambiguous situation involving the choices of trusting or not and of being trustworthy or not, Low scorers are more likely to be Trusting and Trustworthy while High scorers are more likely to be Suspicious and Untrustworthy (exploitatively oriented). These results are obviously concordant with the description of what the F scale was intended to measure and with the conception of the "antidemocratic personality" as advanced in *The Authoritarian Personality* (Adorno et al., 1950).

However, it is not necessary to posit as do the California authors (Adorno et al., 1950) that the characteristics of the authoritarian personality are to be

> Understood as expressions of a particular kind of personality structure within the personality. The most essential feature of this structure is . . . that the conscience or superego is incompletely integrated with the self or ego. . . . There is some reason to believe that a failure in superego internalization is due to weakness in the ego, to its inability to perform the necessary synthesis. . . . (p. 234)

An alternative viewpoint would be that the personality differences between High and Low scorers on the F scale do not necessarily reflect *structural* differences in personality so much as *content* differences in the values that have been internalized as a result of the individual's reaction to his socialization experiences in a particular social milieu, characterized by a given value pattern. This latter viewpoint suggests that Highs and Lows, Trusting people and Suspicious people do not necessary differ in superego integration or in ego weakness (a common synonym for "psychopathology"). As we have indicated more fully elsewhere (Deutsch, 1957), there are characteristic forms of pathological trust (e.g. "gullibility") as well as of pathological suspicion (e.g. "paranoia"). Presumably, the pathologies of trust and suspicion both reflect internal conflict and ego weakness, both of which may be found in individuals who have internalized widely differing values.

SUMMARY

*S*s played an interpersonal game which, in one position, required them to choose between being trusting or suspicious of another and, in a second position, required them to choose between being trustworthy or untrustworthy toward another. There was a striking tendency for *S*s who were trusting to be trustworthy and for *S*s who were suspicious to be untrustworthy. F scale scores correlated significantly with game behavior; *S*s with Low scores tended to be Trusting and Trustworthy while *S*s with High scores tended to be Suspicious and Untrustworthy in their game choices.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harpers, 1950.

CHRISTIE, R., & COOK, P. A guide to published literature relating to the Authoritarian Personality through 1956. *J. Psychol.*, 1958, **45**, 171–99.

DEUTSCH, M. Conditions affecting cooperation. *USN tech. Rep.*, 1957, NONR-285(10). (Mimeo.)

DEUTSCH, M. Trust and suspicion, *J. conflict Resolut.*, 1958, **2**, 265–79.

MEAD, G. H. *Mind, self, and society,* Chicago: Univer. Chicago Press, 1934.

PARSONS, T., & BALES, R. F. *Family, socialization, and interaction process.* Glencoe: Free Press, 1955.

(Received June 19, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 141–143

MOLAR VARIABILITY IN THE MENTALLY DEFECTIVE[1]

PAUL S. SIEGEL
University of Alabama

AND JAMES G. FOSHEE
Sunland Training Center

WE ARE concerned here with variability as it obtains among equally strong molar responses possessing instrumental equivalence. Upon successive presentations of a grossly constant stimulus occasion, variability among these alternatives may be observed to occur as an amalgamation of what Fiske and Rice (1955) have termed Type I and Type II variabilities. A description of the apparatus employed may provide clarification.

S is presented with the task of choosing among four switches. Approximately equal effort is required to execute each movement. All choices accomplish the common end of turning off a light, and it is assumed that *S* brings into the situation roughly equal response tendencies. A set is established through verbal instructions. *S* then responds to successive presentations of the light signal. Variability is expressed in the pattern of choices.

In this situation, a number of conditions may be varied. Within a factorial design, we have examined the effect of intellective impairment, sex, "motivation," and intertrial delay. But, for present purposes, our concern is with intellective impairment. How does the behavior of the mentally retarded differ from that of the "normal"?

METHOD

Subjects and Apparatus

The experiment employed 128 *S*s. Sixty-four were recruited from the Cloverbottom Home, an institution for the mentally defective. Both familial and organic types were represented. This distinction was felt to be unnecessary for the purpose of the study. Gross sensory and motor disturbances were excepted. Chronological ages ranged from 11 to 46 yr. with a median of 28. Half the *S*s were male. Mental ages (Stanford-Binet) ranged from 4–0 to 10–2 yr. with a mean of 6–7 and a *SD* of 15.1 mo. Seventeen of the 64 *S*s were on a mild drug regimen (dilantin, serpasil, or phenobarbital). They were randomly assigned to conditions. To establish the absence of a confounding drug effect, the performance of these *S*s was separately analyzed upon the conclusion of the experiment. Two comparisons were made. The mean variability score (39.6) of this subgroup was compared with the mean of the total group (37.9) and with the mean (36.5) of 17 nondrugged mental defectives chosen at random from the remaining *S*s within each matching condition. In both instances, the difference is readily attributed to chance. It is felt, therefore, that these *S*s were representative in their task performance, and their results have been included in the final analysis.

The other 64 *S*s were drawn from the Peabody Demonstration School and the Tennessee Orphan Home. The Lorge-Thorndike Intelligence Test (nonverbal battery) was utilized to establish normal intelligence. Half the *S*s were male. Chronological ages ranged from 5–10 to 9–8 with a median of 8–4. Mental age equivalents ranged from 5–0 to 12–2 yr. with a mean of 8–8 and *SD* of 23 mo. The designation of a control group for an institutionalized, mentally defective population is necessarily done at the cost of introducing some element of arbitrariness. We had hoped to realize a closer match for mental age, but difficulties arising in the recruitment of a "normal" group dictated otherwise. It should be noted that our "normals" possessed a higher mean mental age and manifested greater dispersion in this characteristic.

In detail, the apparatus consisted of a half-inch plywood board, painted red, upon which were mounted four micro-switches (hereafter, arbitrarily designated from left to right as 1, 2, 3, or 4) and a light (7½ w.). The board was tear drop in shape with a long dimension of 28 in. The diameter of the rounded end was 21 in. The switches were arranged in a semicircle about a central black spot, 5 in. in diameter. *S* was seated in a chair before a small table. The board was then positioned by *S* to provide equal access to all four switches with the forearm resting upon the narrow end of the board. Each switch was of yellow plastic, was mounted atop the board, and possessed ¼-in. shaft which projected about ½-in. above the switch body. Depression of the shaft generated a clearly audible click and served to extinguish the light. The light bulb was mounted behind a translucent pane, 1 in. in diameter, within a small wooden compartment

---

[1] This study was undertaken while the senior author was serving as Visiting Professor at Vanderbilt University in the summer of 1957. In part it was supported by the Research Committee of the University of Alabama. The authors wish to acknowledge the assistance of Mr. Ben Allen in gathering the data.

[2] Fort Myers, Florida.

TABLE 1

MEAN INDEX OF VARIABILITY OBTAINED UNDER VARIOUS CONDITIONS

	High Motivation				Low Motivation			
	(Short Delay)		(Long Delay)		(Short Delay)		(Long Delay)	
	M	*F*	*M*	*F*	*M*	*F*	*M*	*F*
Mentally Defective	37.3	41.5	34.5	39.3	36.0	43.5	35.0	36.0
Normal	28.4	33.5	35.7	31.0	22.5	37.0	20.0	28.3

which extended beyond the tear drop in a position midway between switches two and three.

As one condition, threat of shock was utilized as a "motivater." A zinc electrode was strapped to the nonpreferred hand and the lead was plugged into a nonfunctional GSR apparatus located within *S*'s field of vision.

Instructions to *S* were as follows:

> Please stand here. When this light comes on (*E* turns on the light), I want you to turn it off. Any one of these buttons will turn off the light. Watch me. See, I always start from here (*E* places his hand on the black spot) and when I turn the light off, I put my hand on this spot (*E* operates the switches in a 4-3-2-1 order). Now you try it (*S* is then seated). Move the chair until you are comfortable and can reach all the buttons easily. Let's do a few for practice. Try this one; now this one; now this one; and now this one (*E* directs *S* to use the switches in a 2-4-3-1 order and insures that *S* is centered).
>
> O.K. Do you understand what I want you to do? Remember, when the light comes on, you turn it off. Use *any* button you want to. Always bring your hand back to the spot after pushing a button. Ready?"

When threat of shock was employed, the following instructions were added:

> Turn it off as fast as you can. If you wait too long to turn the light off, you may get a little electric shock. The shock won't hurt but it will give you a funny feeling. Just hurry each time and you will *not* get shocked at all. Ready?

Each *S* was then given 33 successive trials; a trial consisting of light presentation followed by response. Switch choice was recorded by *E*, seated to the left and slightly behind *S*.

The major subgroups, mentally defective and normal, were equally split into male and female subgroups. In turn, these subgroups were split equally into high (threat of shock) and low (no threat) motivation subgroups. Short (2 sec.) and long (30 sec.) intertrial delays provided the basis for a fourth split in the design. Table 1 depicts the design and presents the outcome.

RESULTS

The 33 responses yielded by each *S* were tabulated in a 4 × 4 contingency table. Each successive pair of responses defined an entry. If, for example, on the first four trials, *S* yielded the successive choices, 3-3-1-4, an entry was made in the third column and third row cell, the third column and first row cell, and the first column and fourth row cell. *S*s persisting in the choice of a single switch yielded 32 entries of a single cell. *S*s staying with a two-switch pattern—say, 1-2-1-2-1-2, etc.—yielded 16 entries each in two cells. *S*s consistently following a four-switch pattern—say, 1-2-4-3-1-2-4-3, etc.—yielded eight entries each in four cells. The upper limit of randomness, variability, or response independence would be reflected in two entries each of 16 cells.[3]

From this contingency table, a variability index was computed for each *S*. This index was arrived at by summing the absolute deviations from the expected random value of two for each cell. The deviations were not squared (chi square) because this operation produced a marked skewing of the distributions. The simple sum was taken as an index of variability or response dependence with low scores indicative of high variability or response independence and high scores of low variability or response dependency. The lowest obtained index was 10 (0 was possible), the highest, 60 (yielded by *S*s persisting in the choice of a single switch). This index was found to be distributed in a reasonably normal fashion, and conventional variance analysis was applied.

The mean value obtained from the eight *S*s examined under each set of conditions is reported in Table 1. As a total group, the mentally defective yielded a mean index of 37.9, the normal 29.6. An *F* of 21.34 indicates a high degree of reliability (*P* is far less than .01). We conclude that in the present experimental task, the mentally defective express significantly less variability (or greater response dependency). For the purpose of this study, we may neglect other primary effects. Interaction of intellective impairment with other "dimensions" is of concern, but no interactions were found to be statistically significant.

DISCUSSION

Lewin's early analysis of the feebleminded (1935) permits a kind of intuitive grasp of the present findings. Response variability possesses face validity as an index of "rigidity." This position, like the related one of Kounin (1941a, 1941b), predicts that the more rigid mental defective will show less variability in the present task.

---

[3] Emil W. Menzel pointed out the applicability of this analysis to the present data.

Even a cursory reading of the literature of the past few years reveals that the concept of rigidity has fared badly. As an alternative, the present investigators wish to propose an S-R account. The principle of disinhibition presents attractive possibilities. Admittedly invoked *ad hoc*, this mechanism can "explain" the outcome of this study. It may possess broader scope as a theoretical account of the behavior of the mental defective. Since its implications are testable, its generality can be experimentally assayed.

If we assume that some sort of inhibition is laid down with the execution of a two-choice pattern—say, 1 followed by 3—any condition that serves to disrupt or dissipate this state will facilitate the repetition of that pattern. Variability is decreased, and for a series of choices, the overall index of response dependency is increased. The mental defective, manifesting a kind of "scatter" to his behavior, is commonly described as "distractible." This invites the present interpretation that his lowered variability results from a more pronounced disruption of inhibition, arising out of his lessened capacity to shut out even minor distraction. This view is analogous to that offered by Hovland (1936) to account for an "anomalous" upsurge of the extinction curve on the initial trial following the cessation of reinforcement. Hovland felt that the sudden shift from reinforcement to nonreinforcement constituted a condition sufficient to dispel the "inhibition of reinforcement" accumulated during massed learning trials. We see the mental defective as peculiarly susceptible to such disinhibiting influences, hence given to lowered variability or repetitiousness when held within a task framework.

The inhibitory state assumed here can be variously identified as reactive inhibition (Hull, 1943), as stimulus satiation (Glanzer, 1953), or as central in the sense used by Duncan (1956). It makes little difference. We assume only that responding, perceiving, or, broadly, experiencing generates a condition at least momentarily inhibitory to the repetition of the experience. We assume further that this inhibitory state is susceptible to external influence and that greater vulnerability is a distinctive characteristic of the mentally defective. In the present study, the normal flux of the environment was taken as disinhibiting. As a deduction from our interpretation, the effect of controlled stimulation upon the variability of the mentally defective might be examined.

It would be desirable to generalize the present findings to other tasks, to other controls, and to other indices of variability. The task was a loosely governed one. Selection of the control group was determined largely by convenience. And the index of variability possesses elements of arbitrariness. The present study must be viewed as exploratory and as hypothesis-generating, not as hypothesis-testing.

SUMMARY

Mental defectives and normal children were required to choose among four switches. All choices accomplished the common end of extinguishing a light, and approximately equal effort was involved in the execution of each choice or response. In a series of successive responses, the mental defective manifested greater response dependency or less variability.

This finding was placed within the framework of disinhibition theory. It was reasoned that the lowered variability of the mental defective results from a more pronounced disruption of the inhibition to repeat arising out of a lessened capacity to exclude minor distractions.

REFERENCES

DUNCAN, C. P. On the similarity between reactive inhibition and neural stimulation. *Amer. J. Psychol.*, 1956, **69**, 227–235.

FISKE, D.W., & RICE, LAURA. Intra-individual response variability. *Psychol. Bull.*, 1955, **52**, 217–250.

GLANZER, M. Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60**, 257–268.

HOVLAND, C. I. "Inhibition of reinforcement" and phenomena of experimental extinction. *Proc. Nat. Acad. Sci., Wash.*, 1936, **22**, 430–433.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

KOUNIN, J. S. Experimental studies of rigidity: I. The measure of rigidity in normal and feeble-minded persons. *Charact. Pers.*, 1941, **9**, 251–272. (a)

KOUNIN, J. S. Experimental studies of rigidity: II. The explanatory power of the concept of rigidity as applied to feeble-mindedness. *Charact. Pers.*, 1941, **9**, 273–282. (b)

LEWIN, K. *Dynamic theory of personality*. New York: McGraw-Hill, 1935.

(Received May 14, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 144–145

IDENTIFYING JEWISH NAMES:

DOES PREJUDICE INCREASE ACCURACY?[1]

PAUL F. SECORD AND ELLEN SAUMER
University of Nevada

AN EARLY study by Allport and Kramer (1946) indicated that anti-Semitic persons were able to identify photographs of Jews and non-Jews with a higher degree of accuracy than less prejudiced persons, and that the high prejudice group had a tendency to label more photographs as Jewish than the low group. Allport and Kramer (1946) theorized that those who are prejudiced against a minority group are sensitized to the visible signs of identity of members of such groups. Although their study was supported by Lindzey and Rogolsky (1950), Carter (1948) found no relationship between degree of prejudice and accuracy of identification. Moreover, Elliott and Wittenberg (1955) subsequently presented evidence that the relationship found in former studies was an artifact due to response bias, i.e., the tendency of prejudiced individuals to label photographs as Jewish. Scodel and Austrin (1957), in a study designed to eliminate the effects of response bias upon accuracy, found no relationship between anti-Semitism and accuracy of identification of Jews. Like others before them, they also demonstrated that prejudiced individuals do indeed tend to identify more photographs as Jewish.

These studies employed Jewish and non-Jewish photographs to provide cues to identification of minority group membership. Physiognomy, however, provides only a limited variety of cues to identity. If persons with anti-Semitic attitudes become sensitized to the signs of the identity of Jews, they should also be able to utilize cues other than physiognomic ones. An ever-present and familiar cue to minority group identification is a person's name. The purpose of the present study is to determine whether there are differences between high and low prejudiced groups of gentiles on response bias and accuracy of identification of Jewish and gentile names.

METHOD

A Jewish name was operationally defined as one which is held by a person who states, explicitly or implicitly, that he is a Jew. Fifty Jewish names were taken at random from the roster of approximately 600 delegates to the 1953 American Jewish Conference. Twenty-five additional names were taken at random from a directory and were designated as non-Jewish names. None of these names was the same as any name in the sample of Jewish names. The names were combined into one list[2] which was presented to two groups: college students and United States Army personnel. The former consisted of 67 students obtained from three lower division psychology classes at the University of Nevada, and the latter of 197 soldiers at Fort Ord, California.

Since Elliott and Wittenberg (1955) discovered that under certain circumstances accuracy scores are a function of response bias, two procedures were employed. One of these measured response bias, and the other, accuracy uninfluenced by response bias. *S*s were first asked to designate each name on the list as either Jewish or non-Jewish. The frequency with which they designated names as Jewish represented their response bias. When this procedure had been completed they were given another copy of the list and asked to designate exactly 50 of the 75 names as Jewish. The frequency with which they chose actual Jewish names constituted their accuracy score.

Finally, in order to obtain an index of prejudice, *S*s were given a questionnaire consisting of 26 items selected from the California A-S Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), a measure of anti-Semitism. After completing the questionnaire, *S*s were asked to indicate whether they were gentile or Jewish. Data obtained from those who designated themselves as Jewish, or who failed to follow instructions on any of the forms were discarded. The 65 *S*s discarded for the Army group, a large number, did not vary in any systematic way from the retained sample. The greatest single category of error was in failing to designate exactly 50 names in the accuracy variation—an error occurring in spite of carefully worded instructions requiring *S*s to make two counts. *S*s who checked 49 or 51 names were retained, how-

---

[1] The authors wish to thank Howard H. McFann, who made arrangements for obtaining United States Army personnel as subjects, and Robert D. McDonald, who administered the forms to the military group.

This investigation was supported in part by a grant (M-1892) from the National Institute of Mental Health, Public Health Service.

[2] The 75 names have been deposited as Table A with the American Documentation Institute. Order Document No. 6236 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1
ACCURACY AND RESPONSE BIAS AS A FUNCTION OF PREJUDICE

Prejudice	Accuracy		Response Bias	
	Mean	Sigma	Mean	Sigma
College students				
High ($N = 31$)	38.61	2.4	26.03	8.9
Low ($N = 31$)	38.29	2.7	26.16	8.9
Mean difference	+0.32*		−0.13*	
Army personnel				
High ($N = 66$)	36.00	2.7	33.02	10.0
Low ($N = 66$)	36.56	2.6	27.85	12.0
Mean difference	−0.56*		+5.17*	

* $P > .05$ level.

ever, with *E* adding or deleting a name by a random method. Only five college students were deleted, and three of these were discarded because they were Jewish.

RESULTS

"High" and "low" prejudice groups were formed by designating those above the median A-S score for that sample as "high" and the remainder as "low." Since only nine of the 62 A-S scores for college students were in a prejudiced direction from the "neutral point" of the scale, and slightly more than half of the Army personnel had scores in the prejudiced direction, results were treated separately for the two samples. High and low prejudice groups for each sample were compared on accuracy and response bias as reported in Table 1. No significant differences ($P > .05$) between the high and low prejudice groups were found on either response bias or accuracy. As an additional check, means were also computed for the highest quartile and the lowest quartile. Results were essentially the same as those reported in Table 1, with no appreciable or significant differences occurring.

When names were ranked according to the frequency with which they were designated as being Jewish, it was apparent that many were overwhelmingly chosen as Jewish and many rarely so designated. Since, for statistical reasons, names always or rarely identified as Jewish cannot discriminate as readily between high and low prejudice groups, the high and low groups were compared for accuracy on the 25 names in the middle frequencies. Again, however, there was no significant difference between the two college groups (mean diff. = −0.12), or the Army groups (mean diff. = +0.18).

Finally, the question was raised as to whether the obtained mean frequency of approximately 38 correctly identified Jewish names (36 for Army personnel) was significantly above the theoretical frequency of approximately 33 names which could be obtained by chance. A one-tailed binomial test (Siegel, 1956) comparing these frequencies, indicated that they were not significantly different (college *S*s, $P > .11$; Army *S*s, $P > .26$).

DISCUSSION

Data obtained here are consistent in every way with the conclusion that there is no difference between high and low prejudice groups either in their ability to discriminate Jewish from gentile names or in their tendency to label names as Jewish. Response bias has usually been obtained in previous studies with Jewish photographs; the present results differ largely in failing to obtain response bias for Jewish names. The reason for the failure to obtain a significant degree of response bias is not clear.

The present negative results for accuracy of identification of names are consistent with the fact that properly controlled photographic studies have been uniformly negative. It seems reasonable to conclude that the hypothesis that prejudice increases accuracy of identification of Jews may be safely put to rest, with only a remote possibility that an exhumation will later be called for.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9–39.

CARTER, L. F. The identification of "racial" membership. *J. abnorm. soc. Psychol.*, 1948, **43**, 279–286.

ELLIOTT, D. N., & WITTENBERG, B. H. Accuracy of identification of Jewish and non-Jewish photographs. *J. aborm. soc. Psychol.*, 1955, **51**, 339–341.

LINDZEY, G., & ROGOLSKY, S. Prejudice and identification of minority group membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 37–53.

SCODEL, A., & AUSTRIN, H. The perception of Jewish photographs by non-Jews and Jews. *J. abnorm. soc. Psychol.*, 1957, **54**, 278–280.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

(Received March 30, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 146–147

TAYLOR MAS AND GROUP CONFORMITY PRESSURE

G. L. MANGAN, D. QUARTERMAIN, AND G. M. VAUGHAN
Victoria University, Wellington, New Zealand

EXPERIMENTAL studies of conformity in a situation of group pressure have increased rapidly in number since Asch's (1952) original demonstration. Such studies suggest that the determinants of conforming behavior might appropriately be classified under three headings involving (*a*) the task stimulus (ambiguity, for example), (*b*) the social influence (number of confederates), (*c*) the characteristics of the individual (personality correlates).

In a previous study (Mangan, Quartermain, & Vaughan, 1959) we investigated a variable of the third type, namely the relationship between conformity scores in a setting like Asch's and scores on the Taylor Scale of Manifest Anxiety (Taylor, 1953). The results showed no significant difference in gross amount of yielding between *S*s who scored high (HA) and those who scored low (LA) on the Taylor Scale. However, a secondary analysis of yielding responses in terms of the *S*s' position in the response sequence, which was taken as a measure of the amount of pressure being exerted on the *S*, showed that the HA *S*s conformed less, and the LA *S*s more, as they responded later in the sequence. We suggested that, as pressure increases, the HA *S*s show an increasing amount of hostility and a resulting unwillingness to conform.

In the present research, we attempt to investigate this effect more directly. We hypothesized that when group pressure is severe, *S*s low in Taylor anxiety should give significantly more yielding responses than *S*s high in Taylor anxiety.

METHOD

Material and Subjects

The experimental material consisted of 20 slides, measuring 2¼ in. x 2¼ in., each mounting two clusters of dots, labelled A and B, which were clearly differentiable.

The 12 highest and the 12 lowest scorers on the shortened 28-item Taylor MAS (Taylor, 1953) were selected from a total of 160 undergraduate students as *S*s for this experiment. The mean Taylor score for the LA *S*s was 3.25, for the HA *S*s 18.25. The confederates were 12 specially selected undergraduate students, who underwent a rigorous training procedure before the experiment was begun.

Procedure

The four confederates and the *S* were instructed to be outside the experimental room at a certain time. The confederates approached from different directions and involved the *S* in a discussion about where the experiment in psychology was to take place and when. The experimenter then appeared and ushered the group into the experimental room. The confederates arranged for the *S* to sit at one end of the semicircle of chairs, which were placed immediately in front of the projector and about five yd. from the screen.

The group was given the following instructions:

> This is an experiment in perceptual speed. I am going to show you some slides for a short exposure. On these slides you will see two groups of dots, labelled A and B. I want you to tell me which of the two groups contains the larger number of dots.

The experimenter exposed each slide for 1 sec., and responses were recorded. The order of response was rotated in each trial so that each *S* responded in each of the five positions on four occasions. The predetermined responses of the confederates were coded and embedded in a large block of data written upon the blackboard in the experimental room under the legend "Stage III results, Experiment 15: April 10th." "Please leave" was written across the table.

Four of the 20 trials were considered noncrucial—i.e. in these the confederates gave the correct response without comment. In ten of the remaining 16, the confederates gave incorrect responses without comment. In the remaining six trials, in which the *S* responded third (Trials 8 and 18), fourth (Trials 9 and 19), and fifth (Trials 10 and 20) in the response order, the following procedure was adopted. The first confederate on these six occasions answered correctly in a hesitant manner. He was immediately criticized strongly by the confederate or confederates following him, but preceding the *S* in the response sequence. These gave the incorrect response. The first confederate appeared confused and then conformed. In these trials, therefore, pressure was exerted indirectly on the *S* before he was required to respond. It was assumed that the criticism levelled at the first confederate and his subsequent conforming behavior constituted pressure on the *S* about to respond. To avert suspicion about the object of the experiment it was considered essential to limit the number of these "strong pressure" trials to six.

At the end of the session the *E* disclosed the true nature of the experiment to the *S*.

RESULTS

The 12 HA *S*s gave 36 yielding responses (19%) compared with 64 (33%) for the 12 LA *S*s. This difference in gross amount of yielding, when tested by a Mann-Whitney *U* Test, is significant at better than the .001 level.

The 16 crucial trials were ordered into three categories: the six strong pressure trials, in which there was interchange of comment between the confederates, the three pressure trials immediately

TABLE 1

PERCENT OF SUBJECTS YIELDING IN THE THREE CATEGORIES OF TRIALS

Category of Trial	Percentage Yielding	
	LA	HA
Strong Pressure Trials	25	8
Immediately following Trials	50	28
Pressure Trials	35	23

following the strong pressure trials and the remaining seven pressure trials in which the confederates gave the incorrect responses without comment. Percentages of yielding responses in the three blocks of trials are presented in Table 1.

The differences in yielding percentages in the strong pressure and the immediately following trials are significant for both the LA *S*s ($P < .006$) and the HA *S*s ($P < .04$) by the sign test. None of the other differences is significant, although a comparison of yielding percentages of HA *S*s in the strong pressure and the pressure trials is suggestive ($P < .07$).

DISCUSSION

Judging from Asch's original data (32%) and from those reported in the previous study (33% for the LA *S*s, 27% for the HA *S*s) it is clear that the difference in yielding between the HA and LA *S*s in this experiment is due to the low conformity of the HA *S*s. The LA *S*s conform with about the same frequency as previously reported. The original prediction therefore is strongly supported.

A clear trend is evidenced in Table 1. Both groups, particularly the HA *S*s, show very much lower yielding percentages for the six strong pressure trials, and both groups show significantly more receptivity to pressure in the three immediately following trials, the latter effect being stronger among the LA *S*s.

The reactions of the HA and the LA *S*s to the experimental procedure were strikingly different. The LA *S*s appeared disoriented and confused and often made self-disparaging remarks and asked for instructions, particularly in the strong pressure trials. On the other hand, the HA *S*s seemed somewhat suspicious, occasionally criticized the confederates and showed signs of anger or hostility or both. This was clearly discernible in the post-experimental enquiries. We include a brief description of one such enquiry each, from an HA and an LA *S*.

Q. Did you think there was anything strange about the whole setup?

LA	HA
A. No. I thought after a while that I'd made a mistake and that I was supposed to name the group having the smaller rather than the larger number of dots. [At this point, many LA *S*s asked for instruction.] After that, I was rather confused.	A. Probably; weren't the others playing follow the leader—trying to have me on or something?

Q. Did you think there was any collusion between the *S*s?

A. At one stage I had a vague suspicion, but didn't really know.	A. Generally, yes. At times they sounded very convincing and I wasn't at all sure; why should I change my mind anyway no matter how many think differently.

Q. Did you think there was any collusion between the *E* and the other *S*s?

A. No.	A. No.

The impression gained from these enquiries and from observations made during the course of the sessions is that whereas the LA *S*s tend to question their own judgment more but still to show "normal conformity," under strong pressure conditions the HA *S*s become somewhat hostile and rigid and less amenable to conformity influence. Some general support is thus given to the findings of Janis and Feshbach (1953) that increase in emotional tension reduces the effectiveness of persuasive communications. More specifically, these results substantiate Janis' (1954) original hypothesis that *S*s who show marked neurotic anxiety symptoms as measured on selected items of the MMPI and the United States Army Neuropsychiatric Screening Adjunct are less compliant to persuasion than those *S*s who score lower on these items. It must be pointed out, however, that Janis failed to confirm his earlier hypothesis in a later study (Janis 1955) in which he used a questionnaire derived by Sarason & Gordon (1953) as a measure of neurotic anxiety.

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice Hall, 1952.

JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.

JANIS, I. L. Anxiety indices related to susceptibility to persuasion. *J. abnorm. soc. Psychol.*, 1955, **51**, 663–667.

JANIS, I. L., & FESHBACH, S. The effects of fear arousing communication. *J. abnorm. soc. Psychol.*, 1953, **48**, 78–92.

MANGAN, G. L., QUARTERMAIN, D., & VAUGHAN, G. Relationship between Taylor MAS scores and group conformity. *Percept. mot. Skills*, 1959, **9**, 207–209.

SARASON, S. B., & GORDON, E. M. The Test Anxiety Questionnaire scoring norms. *J. abnorm. soc. psychol.*, 1953, **48**, 447–448.

TAYLOR, J. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received May 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 148–150

PERSONALITY AND NEGRO-WHITE INTELLIGENCE[1]

SHELDON R. ROEN
University of New Hampshire

THE SUPREME COURT in its decision on school integration found that segregation in the public schools implies the inferior status of Negro children thereby retarding their "educational and mental development." This ruling has evoked considerable controversy among psychologists. McGurk (1956), in a widely circulated magazine article purporting to survey "the *only* existing studies that relate to the problem" (p. 96) so as to make "sense" out of the Supreme Court's decision, concluded that "there is something more important, more basic to the race problem than differences in external opportunity" (p. 96). He implied that there was a contradiction between the Supreme Court's decision and the psychological literature. Shuey (1958) in a book reviewing the research on Negro intelligence concludes that the evidence "points to the presence of some native differences between Negroes and whites as determined by intelligence tests" (p. 318). Reactions to this resurgence of racist theory (Klineberg, 1957; Long, 1957; North, 1956) stress the lack of justification for generalizing about race differences, but add no new data.

It seems that a rather important dimension, that of personality as it is manifested in self-perceptions, has been overlooked. A search of the literature reveals that there is evidence: (*a*) that the personalities of Negroes are different from whites (Boyd, 1952; Karon, 1957); (*b*) that personality is related to ability to perform on intelligence tests (Sontag, Baker, & Nelson, 1955; Trumbull, 1953); (*c*) that as infants (Gilliland, 1951) and preschool children (Anastasi & D'Angelo, 1952), Negroes do not score lower on measures of intelligence (these two studies were overlooked by Shuey and by McGurk); and (*d*) that the emergence of personality is closely associated with environment. However, no studies were found that relate the points to one another.

Viewing the issue in this way suggests the proposition that indirect psychological experiences of socioeconomic exclusion, complicated by feelings of lack of historical achievement, and generally more erratic family ties, negatively influences the emerging personality or self-perceptions of Negroes, especially as these relate to their intellectual potentials. Before the proposition can be dealt with directly, the relationship of certain personality traits to intellectual variables in Negroes should be established. The following hypotheses were formulated: (*a*) Negroes would score lower on measures of intelligence than would whites; (*b*) Negroes would score less positively on measures of certain personality traits than would whites; and (*c*) personality variables would correlate more highly with intelligence test scores in Negroes than they would in whites.

METHOD

The personnel files of hundreds of soldiers stationed in an Army post in the Southwest were examined for biographical data and Army Classification Battery (ACB) scores. About 500 were selected for testing and interviews by eliminating those of extremely low ACB scores and those whose biographical data would be hard to duplicate. Personality tests available for administration were the California Test of Personality (Thorpe, Clark, & Tiegs, 1953), the Taylor Manifest Anxiety scale (Taylor, 1953), and the Bernreuter Personality Inventory (Bernreuter, 1935). A plea was made for the soldiers' cooperation and those not desiring to participate were excused. The tests were taken under conditions of anonymity and 98% cooperated.

A measure of test-taking "integrity" was developed by selecting items on the tests that on face value (judgments of three clinicians) appeared to be contradictory. *S*s who responded in a contradictory manner to three or more pairs of items were eliminated from the sample.

Fifty Negro and 50 white soldiers were then closely matched on 10 variables: age, education, occupation of parents, income of parents, geographical area of childhood, army rank, number of years in the service, marital status, urban or rural background, and plans for re-enlistment. Matching was accomplished by categorizing and subcategorizing the variables and then pairing a Negro with a white soldier who most closely resembled him in biographical pattern. A typical example of a matched pair was where both: (*a*) were married and between 25 and 35 years old, (*b*) graduated from high school but did not enter college, (*c*) had fathers who were unskilled laborers, (*d*) indicated that the income of their parents was below $5,000 a year, (*e*) were brought up in the Northeast

---

[1] A modification of this paper was read at the APA convention in Cincinnati, 1959.

The analysis of the data for this research was supported by the Central University Research Fund of the Graduate School at the University of New Hampshire.

TABLE 1

DIFFERENCE BETWEEN MEAN SCORES FOR NEGROES AND WHITES

Measure	Negro		White		*p*
	M	*SD*	*M*	*SD*	
Combined intelligence	91.3	16.5	101.2	19.0	<.005
Pattern analysis	89.2	20.2	103.0	21.6	<.005
Total adjustment	127.7	17.5	129.2	20.6	*NS*
Manifest anxiety	13.8	11.2	12.9	8.3	*NS*
Lack of self-confidence	5.6	83.6	−43.0	67.6	<.005

section of the United States and in cities over 500,000 in population, (*f*) were sergeants, (*g*) were in the army from 6 to 10 years and planned to reenlist. Although this type of matching might tend to mask some important personality differences that might be found in representative Negro and white groups, the desirability of careful matching for the purpose of obtaining comparable correlations between measures of intelligence and personality in the two groups seemed to outweigh the likelihood of obscuring such group differences in personality.

Two measures of intelligence were derived from ACB scores. The first measure, the "combined intelligence measure", was obtained by averaging the scores on three tests that form part of the Army Classification Battery. These tests were found (Montague, Williams, Lubin, & Giesking, 1957) to correlate with the Wechsler-Bellevue as follows: RV (reading and vocabulary) .77, AR (arithmetic reasoning) .78, and PA (pattern analysis) .81. The PA score, since it correlates highest with the Wechsler-Bellevue, was used as a second measure of intelligence.

RESULTS

On both the combined intelligence measure and the Pattern Analysis test, the Negro group scored significantly lower than the white group. The mean Negro score was 91.3 for the combined measure, and 89.2 for pattern analysis. The mean white score was 101.2 for the combined measure and 103.0 for pattern analysis.

Three scores were used to measure personality variables: (*a*) the total adjustment score on the California Test of Personality, (*b*) the Taylor Manifest Anxiety Score, and (*c*) the self-confidence scale score of the Bernreuter Personality Inventory. Only the self-confidence score showed a significant difference between the two groups, with Negroes scoring in the direction of lack of self-confidence.

Correlations between the two measures of intelligence and the three measures of personality were all significant for the Negro group. The highest relationships for the Negro group were obtained between pattern analysis and lack of self-confidence (−.88), between the combined intelligence score and lack of self-confidence (−.79), and between pattern analysis and total adjustment (.73).

TABLE 2

DIFFERENCES BETWEEN CORRELATIONS FOR NEGROES AND WHITES

Personality Measure	Combined Intelligence			Pattern Analysis		
	Negro	White	*p*	Negro	White	*p*
Total Psychological adjustment	.45**	.35**	*NS*	.73**	.22*	<.001
Manifest anxiety	−.55**	−.27*	<.05	−.40**	−.23*	NS
Lack of self-confidence	−.79**	−.50**	<.01	−.88**	−.46**	<.001

* Significant beyond the .05 level.
** Significant beyond the .01 level.

For the white group, significant correlations were obtained between pattern analysis and lack of self-confidence (−.46), between combined intelligence score and lack of confidence (−.50), and between combined intelligence score and total adjustment (.35).

The differences between the correlations for the Negro and white groups on the same correlated variables showed significance beyond the .01 level between correlations with respect to total adjustment and pattern analysis, lack of self-confidence and the combined intelligence score, and lack of self-confidence and pattern analysis. The difference between correlations for anxiety and the combined intelligence score was significant at the .05 level. In each case the correlation associated with the Negro group was larger.

DISCUSSION AND CONCLUSIONS

The personality tests selected were chosen on face value as possibly associated with ability to score on intelligence tests. Of the scales available, psychological adjustment, manifest anxiety, and self-confidence seemed to fit this criterion. High correlations between the measures of intelligence and the personality variables were obtained. That these measures of personality in Negroes were more associated with ability to score on intelligence tests than they were with whites was demonstrated by the significant differences between the correlations for the two groups.

Part of the relationship between the personality measures and intelligence might be explained in terms of the wide spread of intelligence in the sample (intelligence measures ranged from 70 to 139). Taylor (1955) cited evidence of this relationship. However, the range and standard deviations of the measures used are not so different for the

Negro and the white groups as to compel the conclusion that the results can be fully accounted for in this way.

The hypotheses tested were confirmed: the Negro group scored lower on the ACB than did the whites; the Negro group manifested a greater lack of self-confidence than did the whites; and all three measures of personality showed higher correlations with ACB scores for the Negro group than they did for the white group.

Although statements of causality cannot be made from the data presented, the evidence can be seen as warranting further research on the proposition that Negroes as a group, lacking support from pride in significant historical achievement, and developing in an environment of negative experiences, incorporate intellectually defeating personality traits that play a significant role in their ability to score on measures of intelligence.

REFERENCES

ANASTASI, ANNE, & D'ANGELO, RITA Y. A comparison of Negro and white preschool children in language development and Goodenough Draw-a-Man IQ. *J. genet. Psychol.*, 1952, **81,** 147–165.

BERNREUTER, R. G. *Manual for the personality inventory.* Stanford: Stanford Univer. Press, 1935.

BOYD, G. F. The levels of aspiration of white and Negro children in a nonsegregated elementary school. *J. soc. Psychol.*, 1952, **36,** 191–196.

GILLILAND, A. R. Socio-economic status and race as factors in infant intelligence test scores. *Child Develpm.*, 1951, **22,** 271–273.

KARON, B. P. *The Negro personality.* New York: Springer, 1957.

KLINEBERG, O., ET AL. On race and intelligence: A joint statement. *Amer. J. Orthopsychiat.*, 1957, **27,** 420–422.

LONG, H. H. The relative learning capacity of Negroes and whites. *J. Negro Educ.*, 1957, **26,** 121–134.

McGURK, F. C. J. "Psychological tests": A scientist's report on race differences. *U.S. News World Rep.*, 1956, Sept. 21, 92–96.

MONTAGUE, E. K., WILLIAMS, H. L., LUBIN, A., & GIESEKING, C. F. The use of army tests for the assessment of intellectual deficit. *USA res. Rep.*, 1957, WRAIR–52–57.

NORTH, R. D. The intelligence of American Negroes. *Res. Rep. Anti-Defamation League B'nai B'rith*, 1956, **3,** 2–8.

SHUEY, AUDREY, M. *The testing of Negro intelligence.* Lynchburg, Va.: Bell, 1958.

SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA. Personality as a determinant of performance. *Amer. J. Orthopsychiat.*, 1955, **25,** 555–562.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc., Psychol.*, 1953, **48,** 285–295.

TAYLOR, JANET A. The Taylor Manifest Anxiety Scale and intelligence. *J. abnorm. soc. Psychol.*, 1955, **51,** 347.

THORPE, L. P., CLARK, W. W., & TIEGS, E. W. *California Test of Personality.* Los Angeles: California Test Bureau, 1953.

TRUMBULL, R. A study in relationships between factors of personality and intelligence. *J. soc. Psychol.*, 1953, **38,** 161–173.

(Early publication received December 2, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 151-154

PATTERNS OF PERMISSIVENESS AMONG PRELITERATE PEOPLES

E. TERRY PROTHRO
American University of Beirut

IS PERMISSIVENESS in child rearing a general characteristic of maternal behavior—a dimension, a trait? Or does permissiveness of mothers vary with the type of child behavior? Studies of American mothers do not seem to yield consistent answers on this point. Sewell (1955) interviewed a sample of middle-class rural Wisconsin mothers and found little relationship among various kinds of permissiveness. Sears (1957) interviewed working class and middle-class mothers in two New England suburbs, and found a general permissiveness factor. Such inconsistency of results in studies of specific communities seems fairly typical of recent American surveys of child rearing practices. Reports on practices in Illinois (Davis, 1952), Massachusetts (Maccoby & Gibbs; 1954), Oregon (Littman, Moore, & Pierce-Jones, 1957), and California (White, 1957) differ to such a degree that most students of maternal behavior would concede with Havighurst and Davis (1955, p. 441) that "we should not attempt to generalize concerning child-rearing . . . from a sample in one part of the country, even if it is a representative sample."

As Whiting has pointed out (Whiting, 1953), contradictory results in studies of individual communities often may be resolved by utilizing cross-cultural data. Such data reduce the possibility of negative findings attributable to restricted range of behavior, or of positive findings attributable to temporary or local factors. These considerations highlight the importance of the study by Whiting and Child (1953) of the child rearing norms of cultures from all parts of the world. From this study they concluded (p. 117) that among nonliterate peoples "the [child rearing] practices of a society for one system of behavior are almost entirely independent of its practices with respect to other systems of behavior." Cultural norms of child rearing, then, do not differ with respect to permissiveness in general, but differ with respect to permissiveness for sex behavior, permissiveness for dependent behavior, etc. This assertion is, as they pointed out, one of considerable theoretical and practical importance. It is with that assertion, and the data on which it is based, that the present study is concerned.

SOURCE OF DATA

In their cross-cultural study, Whiting and Child (1953) gathered data on child rearing in 75 cultures by utilizing the Human Relations Area Files and some other "pertinent articles or monographs which had not yet been entered in the file." This information was placed in the hands of experienced judges, who assigned ratings (whenever possible) to each culture with respect to its indulgence or permissiveness for five "systems of behavior"—oral, anal, sexual, dependence, and aggression. For each system, separate judgments were made of initial satisfaction, age at serious socialization, and amount of socialization anxiety. Finally, each culture was rated (whenever possible) with respect to certain customs relating to illness. All of these ratings are reported in the appendix of *Child Training and Personality* (Whiting & Child, 1953, pp. 338–346).

Although they were interested in the interrelations of the child rearing practices, Whiting and Child did not carry out a factor analysis of their data. For the purposes of this study, it was decided to re-examine their data by the use of factor analysis. Two analyses were carried out. Thurstone's centroid method was used, and axes were rotated graphically. The first analysis was based on ratings assigned those 22 cultures on which fairly complete information was available. The following cultures were included because they had all been rated on 30 of the practices: Chagga, Comanche, Dahomeans, Hopi, Kwakiutl, Kwoma, Lepcha, Lesu, Manus, Marquesans, Navaho, Ontong-Javanese, Papago, Pukapukans, Siriono, Tanala, Tenino, Teton, Western Apache, and Wogeo. After the initial analysis had been completed, and it had been established that there were several dimensions of child rearing, a factor analysis was made of data from 36 cultures to determine whether the same dimensions would emerge with a larger sample. In the second analysis, the 22 cultures of the first analysis were included, and added to them were 14 cultures on which ratings were available for 12 child rearing practices and 7 illness practices. By reducing the number of practices included in the analysis it was possible to include a larger number of cultures. The additional cultures were: Ainu, Arapesh, Azande, Bena, Chamorro, Chenchu, Chiricahua, Kurtatchi, Kutenai, Maori, Samoans, Thonga, Trobrianders, and Yakut.

RESULTS

Rotated factor loadings for each of the analyses are shown in Table 1. It can be seen that the three factors emerging from the first analysis correspond quite closely to those of the second analysis.

The factor identified as *F2′* in each of the

TABLE 1
ROTATED FACTOR LOADINGS

Variable	First Analysis (22 Cultures)			Second Analysis (36 Cultures)		
	$F1'$	$F2'$	$F3'$	$F1'$	$F2'$	$F3'$
Initial Indulgence						
A Oral	−11	−65	−08	−08	−63	−28
B Anal	01	25	−78	−22	32	−71
C Sexual	00	61	−21	−42	40	−20
D Dependence	−07	−58	11	13	−49	00
E Aggression	−56	17	−22	−52	−05	−24
Socialization Anxiety						
G Oral	44	43	−16	27	54	00
H Anal	02	−11	75	07	−04	85
I Sexual	07	−73	12	51	−53	17
J Dependence	33	24	−50	17	42	−40
K Aggression	48	−15	16	66	09	14
Age at Socialization						
M Weaning	11	−55	10	−05	−46	−27
N Toilet Training	01	−22	−63	—	—	—
Q Independence	−18	−14	54	20	−47	51
Illness Explanations						
a Oral	49	48	−13	08	52	−21
b Anal	68	−02	29	54	02	07
c Sexual	44	−10	−22	42	−06	−11
d Dependence	03	73	01	−04	57	06
e Aggression	71	01	−34	52	32	−26
Performance Therapy						
f Oral	39	−09	18	—	—	—
g Anal	55	−04	14	—	—	—
i Dependence	37	−12	06	—	—	—
j Aggression	41	−06	15	—	—	—
Avoidance Therapy						
k Oral	93	−10	09	—	—	—
l Anal	71	−04	07	—	—	—
m Sexual	36	−29	−15	—	—	—
n Dependence	12	02	−45	—	—	—
o Aggression	30	−07	15	—	—	—
Fear of						
p Human beings	38	−21	16	58	−07	24
q Spirits	−34	34	12	−21	25	18

analyses could be called an *orality–sexuality* factor. The culture high on this factor characteristically indulges the infant's oral needs, weans late, evokes little anxiety over weaning, indulges the infant's expressions of dependency, but also is nonpermissive of infantile sexuality and produces considerable anxiety over sex training. Cultures at the other extreme are permissive regarding sex but nonindulgent of the infant's oral and dependency needs. They are also characterized by "oral explanations of illness" and "dependency explanations of illness." That is, these cultures typically explain illness as caused by ingesting something, by verbal spells (oral) or by the soul leaving the body, or as an evil spirit entering the body (dependency).

The factor called $F3'$ in both analyses could be named an *independence–anality factor*. The culture high on this factor begins toilet training early and evokes much anxiety over the training, but it begins serious independence training comparatively late. Cultures at the other extreme are indulgent about the infant's anal activity and late in toilet training, but they begin serious independence training early and produce much anxiety in the process.

The third factor, $F1'$ in these analyses, might be called an *aggression–hypochondria factor*. Almost all of the practices relating to illness had loadings on this factor, so it was called "hypochondria" to emphasize the generalized concern with illness. The only child rearing practices with substantial loadings on this factor in both analyses were the two pertaining to permissiveness for aggression. Cultures high on this factor permitted little aggression in small children, produced anxiety in the process of socializing aggression, and were generally high in the use of various therapies and explanations for illness. Treatment of infantile sexuality had a significant loading on this factor in the second analysis, but no loading at all in the first analysis.

DISCUSSION

The child rearing norms of these cultures cannot be described as permissive or nonpermissive. Rather, they must be described as more or less permissive on orality-sexuality, on anality-independence, and on aggression. It is not even accurate to refer to the "permissiveness for dependency" in a given culture, for dependence is not unifactorial. Permissiveness for dependency behavior in the nursing infant is not correlated with permissiveness for dependency at the time of serious socialization, or the degree of anxiety produced by such independence training. This dual nature of dependency was noted by Whiting and Child (1953, p. 109).

While it cannot be asserted that permissiveness is wholly generalized, neither does it seem correct to say that permissiveness is specific to the system of behavior–oral, sexual, anal, dependence, or aggression. There are among these systems definite clusters or patterns in the cultural norms (cf. Littman & Moore, 1957, p. 703). The assertion of Whiting and Child that "there is no consistent relation between the way one system of behavior is socialized and the way any other system of behavior is socialized" (Whiting & Child, 1953, pp. 115–116) overemphasizes the specificity of norms. They computed intercorrelations among ratings on the five systems for initial indulgence, and repeated this for severity of socialization and age of socialization. Because ratings for many cultures were incomplete, the maximum number of cultures used in computing these correlations was 36. No correlations were computed between ratings of one system on initial indulgence and ratings of other systems on severity of socialization or age at socialization. The "no consistent relation" conclusion was based on the observation that significant correlations among those computed were few, and that positive and negative correlations were about equally prevalent. Such observations do not seem to justify their conclusion. Patterns in the data could have been found by utilizing other techniques.

The bipolarity of the *oral-sexual* factor could be interpreted in any one of a number of ways. The psychoanalytically oriented might argue that the children, and by extrapolation the culture, are fixated at either the oral or the phallic stage of psychosexual development. Others might approach the problem from the standpoint of the adult and say that the mother finds nursing and sex to be alternative types of gratification. In a culture where sex activity is encouraged and accepted, the mother may find prolonged nursing a sex handicap. Thus Linton (1939, pp. 164–165) describes Marquesan mothers who disliked nursing because of a fear that their breasts, considered important to sex play in that culture, might be "spoiled." The converse of this situation seems to exist among some Arab mothers, for whom sex is shameful and frightening, but who confess blushingly that they find nursing their child to be pleasurable (Ammar, 1954, p. 190; Wakim, 1956, p. 50).

The opposition between toilet training and independence training in the *independence-anal* factor is not difficult to understand when it is recognized that much of "independence training" consists of leaving a child to his own resources (Whitney & Child, 1953, p. 97). If the two-and-a-half year old child is allowed considerable freedom in developing his defecation habits, he may well be required at three-and-a-half to develop his own habits of play and self-protection. There is also a possibility that *ecological variables* play a part in opposing "anal" and "independence" indulgence. In a Lebanese mountain-village surrounded on three sides by canyons, and covered in winter with deep snow, a mother may be compelled to restrict the outdoor activity of her three year old. At the same time, the confined quarters would inhibit a casual attitude toward toilet training. The environment would compel parents to delay or ignore independence training and to inhibit "anal indulgence." Beduoins, on the other hand, live more in open country, and are more mobile, so that sanitation is less of a problem. At the same time, it is more feasible for the mother to let the child shift for himself outside the family tent. Their environment would encourage late toilet training and early independence training. Hence, the physical environment may in some instances contribute to the opposition between toilet training and independence training.

Although the *aggression-hypochondria* factor seems to be the same in both analyses, there are two reasons for caution in asserting the presence of such a factor. First, therapy ratings were not included in the second analysis. More important, however, is the very real possibility that the interrelations between the various illness practices may have been an artifact of the scoring procedure. The judges who rated the cultures were instructed to give ratings of plus when the ethnographic material on a culture mentioned a certain practice and minus when there was mention of the absence of a practice *or no mention of the practice at all.* If one ethnographer, then, noted carefully a wide variety of beliefs and practices regarding illness, and another noted only the most striking ones the second culture would be scored as lacking practices found in the first. Whiting (1954 p. 529) has noted that the procedure may yield spurious correlations, and there is a definite possibility that such spurious correlations could have affected the results of our analysis. The "hypochondria" factor might represent ethnographer's generalized concern and lack of concern with illness rather than the culture's concern or lack of it.

SUMMARY

In order to ascertain the generality of norms of permissiveness in child rearing, data published by Whiting and Child (1953) on primitive cultures from all parts of the world were examined by the use of factor analysis. It was found that there were three relatively unrelated dimensions of permissiveness: orality-sexuality, independence-anality, and aggression. Sexual permissiveness was associated with nonpermissiveness on oral matters, and oral permissiveness with nonpermissiveness on sex. Similarly, permissiveness in independence training was negatively related to permissiveness in toilet training. There was some evidence for a general hypochondria factor as a part of the aggression factor. It was pointed out that the patterning of child rearing norms might be explained in terms of conscious motives of parents, ecological variables, or even the concerns of individual ethnographers.

REFERENCES

AMMAR, H. *Growing up in an Egyptian village.* London: Routledge & Kegan Paul, 1954.

DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child rearing. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley, (Eds.), *Readings in social psychology.* New York: Holt, 1952. Pp. 539–550.

HAVIGHURST, R. J., & DAVIS, A. A comparison of the Chicago and Harvard studies of social class differences in child rearing. *Amer. sociol. Rev.*, 1955, **20,** 438–442.

LINTON, R. Marquesan culture. In A. Kardiner (Ed.), *The individual and his society.* New York: Columbia Univer. Press, 1939. Pp. 137–196.

LITTMAN, R. A., MOORE, R. C. A., & PIERCE-JONES, J. Social class differences in child rearing: A third community for comparison with Chicago and Newton. *Amer. sociol. Rev.*, 1957, **22,** 694–704.

MACCOBY, ELEANOR E., GIBBS, PATRICIA K., ET AL. Methods of child rearing in two social classes. In W. E. Martin & C. B. Stendler (Eds.), *Readings in child development.* New York: Harcourt, 1954. Pp. 380–396.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* Evanston: Row, Peterson, 1957.

SEWELL, W. H., MUSSEN, P. H., & HARRIS, C. W. Relationships among child training practices. *Amer. sociol. Rev.*, 1955, **20,** 137–148.

WAKIM, SUAD G. Child care in Mieh-Mieh. Unpublished master's thesis, American Univer. Beirut, Lebanon, 1956.

WHITE, MARTHA S. Social class, child rearing practices, and child behavior. *Amer. sociol. Rev.*, 1957, **22,** 704–712.

WHITING, J. W. M. The cross-cultural method. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954. Pp. 523–531.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality.* New Haven: Yale Univer. Press, 1953.

(Received February 5, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 155-158

CHARACTERISTICS OF UNTESTABLE CHRONIC SCHIZOPHRENICS[1]

HAROLD WILENSKY AND LEONARD SOLOMON[2]
Franklin D. Roosevelt Veterans Administration Hospital, Montrose, New York

RESEARCH studies in schizophrenia which employ clinical tests or laboratory behavior as dependent variables invariably report that some proportion of the *Ss* are untestable or uncooperative. It is tacitly assumed that conclusions drawn from such studies will be delimited to the population of testable patients. Unfortunately, however, the results of such studies generate clinical theory which tends, by and large, to neglect the limitations imposed by the particular character of the testable sample.

It seems desirable to establish criteria for defining "untestability" and to determine which factors are likely to affect this attribute. Moreover, the measure of testability may be a variable worthy of study in its own right in that there is evidence to indicate that it has significant behavioral correlates. Mednick and Lindsley (1958) reported that the rate of response under operant conditioning was related to testability and could have implications for the patient's responsiveness to the environment.

Zubin, Windle, and Hamwi (1953) reported that 25% of the chronic schizophrenics used in the Columbia-Greystone project were untestable on more than one third of a test battery including cognitive (IQ), perceptual, and motor components. Unfortunately, the untestability rates for specific test components were not indicated. King (1954), using relatively simple motor tasks, reported the rate of untestability to be as low as 10% for his chronic schizophrenic sample (median length of hospitalization of 9.1 years). Interestingly enough, Mednick and Lindsley (1958), with an extremely chronic sample (median length of hospitalization of 16 years) on a battery of Wechsler-Bellevue and Rorschach, reported that 55% were untestable. It would appear that the proportion of untestable patients obtained is determined by the characteristics of the patients, task complexity, and social situational factors.

The present study was essentially exploratory in nature and sought to examine the factors producing untestability, to estimate the stability of such behavior, and to compare the differences between testable and untestable patients in order to make explicit the nature of the limitations to be placed upon theoretical generalizations.

PROCEDURE

The sample consisted of the entire population of two wards of 101 chronic male schizophrenic patients. The patients ranged in age from 26 to 70 years with periods of continuous hospitalization ranging from one to 29 years.

Attempts were made to administer individually a Rorschach and an abbreviated WAIS[3] to each patient on three separate occasions, one month apart. The patients were randomly assigned to an experienced clinical psychologist or to relatively inexperienced psychology trainees for all three testing occasions. The tests were administered in a private office adjacent to the ward dayroom. If a patient refused to come to the office or refused to respond to the test questions, he was encouraged to participate by suggesting that the examiner was interested in his progress and that all patients on his ward were being examined.

A number of demographic indices were derived from case history material: age, race, place of birth, education, and length of hospitalization. Indices of current performance were obtained from a variety of sources and by different techniques: diagnosis, mental health ratings, rates of social interaction, and tapping test performance. They are listed and operationally defined below:

Diagnosis. This classification was derived from a mental status examination administered by two psychiatric residents.[4] The three diagnostic categories employed were hebephrenic, paranoid, and other (catatonic, mixed, or simple).

Mental health ratings. Three social psychology trainees, functioning as ward observers, rated each patient on a 6-point scale which focused upon the patient's ability to assume independent responsibility in caring for himself. The scale ranged from ability to live outside the hospital and work (score one) to requiring maximum supervision for his daily activities in the institution (score six). These ratings were made solely on the basis of observed patient behavior on the ward. The interobserver reliability coefficients for the three possible pairs were .73, .77, and .82.

---

[1] Part of this paper was read at the Eastern Psychological Association meetings at Atlantic City, New Jersey, April, 1959.

[2] The authors would like to acknowledge the help given by Alfred Drelich in the statistical analysis of the data. We are also indebted to the Social Psychology Unit for the use of its social interaction data and health ratings, and to the psychology trainees who participated in the repeated testing of patients.

[3] Five subtests were selected, representing each of five factors extracted from a factor analysis of the WAIS by Cohen (1957). They were Information, Digit Span, Picture Completion, Block Design, and Digit Symbol.

[4] Joseph Dorsey and Ralph Baker made the psychiatric diagnoses based upon mental status examinations. Each psychiatrist independently examined approximately half of the cases.

Rates of social interaction. The social psychologists, acting as ward observers, recorded the frequency of patient to patient interactions on the ward over a 3-month period. There were 55 one-hour periods of observation made at random times throughout the day. A measure of the interaction rate for each patient was based upon the following ratio: number of patients' social contacts/number of observation periods in which patient was present.[5]

Tapping test. About 6 months after the administration of the WAIS and Rorschach, the patients received a simple motor test of finger tapping. Conditions differed, however, in that the tapping test was given on the ward, whereas the clinical tests were administered in a private office.

RESULTS

The first testing occasion (Rorschach and WAIS) was used to define testability. Fourteen patients refused to enter the examiner's office despite the encouragement by the psychologist. These patients made clear their intentions not to take tests, and were labeled *untestable refused.* Another 13 patients entered the examiner's office but could not be tested. These patients did not refuse explicitly. Most of them simply replied "I don't know," or made no response to questions. Some were incoherent or irrelevant in their responses. Because they appeared incapable of dealing with the tasks, these patients were labeled as *untestable confused.* The rationale for this classification is based on the assumption of a motivational difference between the confused and refused untestable patients. The confused untestables seemed willing but unable to produce scorable records, whereas the untestable refused group appeared unwilling to perform. Obviously, no judgment could be made as to the ability of the latter group to respond to tests.

Of the remaining patients, 64 gave adequate WAIS and Rorschach protocols. Ten produced only one scorable test (either WAIS or Rorschach). These patients were classified as *testable.* In the subsequent data analyses, the testable, untestable confused, and untestable refused groups were compared with regard to situational, demographic, and current performance variables.

TABLE 1

CHARACTERISTICS AND NUMBER OF PATIENTS IN THE TESTABILITY TYPES

Variable	Testable	Untestable Confused	Untestable Refused	*P* Value
Examiner experience:				
Trainee	36	7	7	*ns*
Staff	38	6	7	
Birthplace:				
Native	69	7	12	<.05[a]
Foreign born	5	6	2	
Diagnosis:				
Paranoid	25	1	10	<.005[b]
Hebephrenic	37	6	3	
Other	12	6	1	
Total	74	13	14	
Central Tendencies:				
Median age	39.5	42.0	46.5	<.02[c]
Median age 1st hospitalization	26.2	26.0	31.0	<.10[c]
Median years continuous hospitalization	9.8	14.5	10.5	<.05[c]
Median education grade	10.0	8.5	8.0	<.10[c]
Median social interaction rate	.063	.033	.050	<.10[c]
Mean mental health rating	3.75	4.88	4.13	d

[a] Fisher's exact test.

[b] Although the requirements of expected cell frequencies of 5 for using chi square were not met, the p value obtained suggests that the null hypothesis may be rejected.

[c] The Kruskal-Wallis test of differences in rank order was used for comparisons among groups.

[d] The untestable confused group significantly differed in mean and in variance from the testable group ($t = 5.12$; $F = 2.90$) and from the untestable refused group ($t = 2.19$; $F = 2.67$). No significant differences were found between the untestable refused and the testable group ($t = 1.24$; $F = 1.08$).

Demographic Variables

Race. No differences in testability related to race were found. Of the fifteen Negroes in the total sample, one refused and one was classified as confused.

Place of birth. Foreign born patients appeared with greater frequency in the untestable confused group. It was likely that language difficulty contributed to the apparent confusion in this group of untestable patients. (See Table 1.)

Age and hospitalization. Age, age at onset of illness, length of continuous hospitalization, and education yielded skewed distributions which were analyzed by means of the Kruskal-Wallis test of differences in rank order (Senders, 1958). With regard to age, the untestable refused group emerged as the oldest. Moreover, the data suggest that their onset of illness may have taken place at a later age. The median period of continuous hospitalization for the refused cases was relatively low. By contrast, the untestable confused group was closer to the median of the testable sample in both chronological age and age of onset of illness. Their period of continuous hospitalization, however, was the longest. In fact, the majority of patients in this confused group have never been separated from a hospital since their first admission. Had the samples been larger, both untestable groups might have shown less education than the testable group. (See Table 1 for medians and *p* values.)

---

[5] In order to test the interobserver reliability of the interaction rate, two observers made independent judgments with respect to the type and role of a patient's participation for 65 interactions. Using an index of reliability for nominal categories known as Kappa (Cohen, 1960) which represents the proportion of observer agreement after chance agreement is removed, we obtained $K = .71$, significant at. <01 level.

Current Performance and Behavior

Diagnosis. Inspection of the distribution of diagnoses revealed that the untestable refused group contained a greater number of paranoid schizophrenics. The untestable confused group contained fewer paranoid patients. (See Table 1 for diagnostic distributions.)

Mental health ratings

Bartlett's test of homogeneity of variance in mental health ratings among the three groups yielded a significant chi square (38.91). Since the assumptions underlying analysis of variance could not be met, the three groups were compared with respect to mental health ratings by means of *t* tests. The poorest ratings occurred in the untestable confused group, which also was less variable than either of the other two samples. No significant differences were found between the untestable refused and the testable group in either mean tendency or variability. The variances for the untestable confused, untestable refused, and the testable groups were .39, 1.06, and 1.15 respectively. (See Table 1 for mean health ratings.)

Rates of social interaction. With respect to social interaction, the findings, although not significant, suggested trends parallel to those observed in the mental health ratings. The untestable confused group seemed to manifest somewhat lower rates of interaction as well as a restricted range when compared with the other two groups. The entire confused group was a bit below the interaction rate median for the entire group. (See Table 1 for median rates.)

Miscellaneous Variables

Examiner experience. The effects of examiner experience did not result in any significant differences in testability. The numbers of patients in each of the testability categories for the staff and trainee psychologists are presented in Table 1.

Consistency. The consistency of testability was determined by repeating the WAIS and Rorschach examinations on two more occasions. Among initially untestable patients, approximately 90% continued to be untestable on clinical tests. Of the initially testable sample, 9 patients refused one or more later examinations. Thus, there was a high degree of consistency in testability. Eighty-eight per cent of the patients were consistent in their accepting or rejecting tests on all three occasions.

Tapping test. The tapping test performance was not included as a criterion of testability because of the different time of administration and examination conditions. It was of interest to note that the percentage of untestable patients dropped to 9% with this simple motor task, which was administered in full view of other patients on their ward. Approximately two-thirds of the untestable patients accepted and performed the tapping test.

DISCUSSION

The results suggest that untestable patients do not compose a homogeneous category. Untestable patients can meaningfully be divided into at least two major subgroups according to *level of ability* in test performance and *motivation* to accept the task. Of the total untestable sample, 13 were cooperative but apparently confused, and 14 explicitly refused to be tested.

Significant differences among the groups were demonstrated for a number of variables such as diagnosis, age, years of continuous hospitalization, and mental health ratings. Although social interaction and age at onset did not attain significance, the pattern of differences was similar to the other variabiles which essentially represented aspects of a more general concept of mental health.

The untestable confused group appeared to be more disturbed on the basis of a variety of criteria: (*a*) their long length of continuous hospitalization, (*b*) their low health ratings, and (*c*) possibly their degree of social withdrawal. Even with the simple finger tapping task, one-third of this group was unable to respond adequately. The characterization of untestable patients made by Zubin et al. (1953) that they are "essentially deteriorated" with poor prognoses would appear to be specifically applicable only to the confused category.

In contrast to this, those patients who explicitly refused to submit to testing were indistinguishable from the testable patients in terms of degree of health, length of continuous hospitalization, or rates of social activity. The untestable refused group differed from the over-all sample in that they were the oldest group, and they were likely to be diagnosed as paranoid. Thus, although this category of patients was untestable, they did not reflect the descriptive image offered by either Zubin et al. (1953) or Mednick and Lindsley

(1958). They could not be considered any more deteriorated or unresponsive to their environment than the testable sample.

Twelve per cent of the sample exhibited fluctuations in type of test performance when the first testing was compared to either the second or third occasion. In effect, the total sample of untestable patients increases by 9% if the entire three testing occasions are used as a criterion. For those investigators who employ a repeated measures design, there are further limitations imposed by this increase in the untestable sample.

On the basis of incidental data, it may be suspected that for different tasks or for different test conditions, testability rates may be substantially altered. For example, the use of simple operant tasks or public testing situations may have a salutary effect on the testability rate. Paranoid patients in particular may be more likely to refuse to enter a private office than to accept tests under visibly safe public conditions.

It is clear from the data that the testable sample, on any test occasion, represents a selectively biased sample of the entire male VA chronic schizophrenic population.

SUMMARY

Attempts were made to examine the entire population of two wards of chronic schizophrenic patients with an individually administered Rorschach and an abbreviated WAIS. A finger tapping test was administered at a later date on the ward. Patients were classified into three groups according to their willingness to accept testing and ability to perform. Twenty-seven per cent of the patients were untestable. The untestable chronic male schizophrenic patients seemed to represent a population qualitatively different from the testable population.

The present investigation therefore affirms the necessity for widening research efforts to include a more articulate description of the "silent sample," untestable schizophrenics.

REFERENCES

Cohen, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, **21**, 283–290.

Cohen, J. A coefficient of agreement for nominal scales. *Educ. psychol. Measmt.*, 1960, **20**, 37–46.

King, H. E. *Psychomotor aspects of mental disease.* Cambridge, Mass.: Harvard Univer. Press, 1954.

Mednick, Martha T., & Lindsley, O. W. Some clinical correlates of operant behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 13–16.

Senders, Virginia L. *Measurement and statistics.* New York: Oxford Univer. Press, 1958.

Zubin, J., Windle, C., & Hamwi, Violet. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, **21**, 342–355.

(Received April 30, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 159-167

THE PROCESS OF COGNITIVE TUNING IN COMMUNICATION[1]

ROBERT B. ZAJONC
University of Michigan

CONSIDERABLE evidence is now available on selective and directive effects in perception and cognition. The earliest explanation of these effects was advanced by the Würzburg School in their postulation of a determining tendency that arises from the subject's (*S*'s) task assignment (*Aufgabe*), and acts as a steering mechanism in information-seeking behavior (Ach, 1905; Lewin, 1922a, 1922b; Watt, 1905). The specific effects of determining tendencies were shown to be intimately related to the nature of the *Aufgaben* (Chapman, 1932; Kulpe, 1904).

The effects of *Aufgaben* are readily detected in communication. Each person is daily a source and a destination of numerous messages, and strong selective and directive effects should result when the individual's role in the communication process is manipulated. Hovland and his associates (Hovland, Janis, & Kelley, 1953) have shown in extensive studies that the characteristics of the communicator and the circumstances of the communication influence the attitudinal and cognitive impact of communications. Zimmerman and Bauer (1956), on the other hand, have demonstrated that communicators themselves are influenced by their audiences. The purpose of this study is to examine the nature of cognitive structures that are activated or "tuned in" when persons enter into communication with others.

For the purposes of this research a set of formal properties were defined for the description and measurement of cognitive structures. Since it is generally agreed that cognitive structures represent organized systems whose nature depends on the various interrelations among their components (Asch, 1946; French, 1947), the main emphasis was placed on morphological description. The properties defined are to some extent based on Lewin's (1951) analysis of the life space and on French's (1947) work on sentiments. Two experiments were conducted in which measures of the cognitive dimensions were obtained and the effects of "tuning" to receive and transmit information were analyzed.

PROPERTIES OF COGNITIVE STRUCTURES[2]

It is assumed that objects and events are perceived and discriminated on the basis of psychological dimensions. By a psychological dimension is understood the organism's capacity to respond to stimuli in such a way that, given a set of stimuli and a set of responses made to them, the stimuli and responses form two ordered sets with a determinate correspondence between the elements of each set. From the person's responses one can infer the values of which psychological dimensions consist. An act of perceiving or discriminating may be thought of in terms of "projecting" the given stimulus on some psychological dimensions, and thereby attributing to it values taken from these dimensions. The values thus assigned to stimuli constitute the elements of cognitive structures and are called attributes. They are the qualities or characteristics that individuals "assign" to objects and events.

An attribute, then, is a concept (or a category) in its simplest form, since a value from a given psychological dimension generally characterizes an entire class of objects. More complex concepts are formed, as Bruner, Goodnow, and Austin (1956) suggest, by conjunctive and disjunctive combinations of attributes. The larger and the more inclusive the set of objects (class) denoted by a concept the more general is the concept. For instance, all sorts of objects are "black," "animal," or "beautiful." Only a few are "76 ft. $4\frac{5}{32}$ ins. long."

A person has at his disposal a large number of concepts or attributes by means of which he can identify and discriminate objects and events in his environment. The set of all such attributes is called his cognitive universe. Let a cognitive structure then be defined as an organized subset of the given cognitive universe in terms of which the individual identifies and discriminates a particular object or event. The morphological properties of cognitive structures describe various relationships among attributes. Those defined here are differentiation, complexity, unity, and organization.

Degree of differentiation. Given a universe of let us say m alternative attributes by means of which a given person is capable of identifying and discriminating objects and events, he can construct a maximum of 2^m distinct cognitive structures. The number (n) of attributes constituting a given cognitive structure is taken to reflect its degree of differentiation.

---

[1] Based upon dissertation in Social Psychology, University of Michigan, in 1954, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to thank Dorwin Cartwright and Daniel Katz for their advice and suggestions in carrying out this study. This investigation was supported in part by a grant to the Research Center for Group Dynamics from the Rockefeller Foundation.

[2] For a more detailed exposition of the descriptive model see Zajonc (1954).

Degree of complexity. The attributes constituting a given cognitive structure may come from a single class or category of discriminanda, or they may represent many categories. For instance, a painting may be perceived exclusively in terms of its objective qualities, such as size, subject matter, color, type of frame, etc. Or, it may be looked upon partly in terms of its objective qualities, partly in terms of its formal properties (period, style, symmetry, harmony, etc.), and partly in terms of its impact on the subjective experience of the viewer. Such attribute groupings may be further subdivided into smaller classes, and the extent of such subdivision may be used in defining the complexity of cognitive structures.

Let us denote the level of inclusion of a given class by r, such that if the class K_i does not include another subclass, $r = 1$; when K_i includes some subclasses, which in turn do not include other subdivisions, $r = 2$; when K_i includes subclasses which themselves contain other indivisible subclasses, $r = 3$, etc. The degree of complexity of a given cognitive structure may be obtained by weighting each attribute by the level of inclusion of its class membership and summing the weighted scores thus obtained. Thus:

$$\text{com} = \sum_{r=1}^{n} r n_r$$

where n_r is the number of attributes in the rth level of inclusion.

Degree of unity. The structural components of a whole may depend on each other to a greater or lesser extent. The more attributes depend on each other the more the cognitive structure is unified. If we define the dependence of the attribute A_i on the attribute A_j as equal to 1 when a change in A_j produces a change in A_i, and as equal to 0 when change in A_j does not produce a change in A_i, then a dependency matrix can be constructed for all attributes of a given cognitive structure, and the total dependency of each attribute may easily be obtained by summing the entries in the appropriate row. To compare the unity of structures of different degrees of differentiation the measure of unity must be normalized. Given a structure with n attributes the maximum dependency of a given attribute, $\text{dep}(A_i)^{\max}$, is equal to $n - 1$, and the maximum sum of dependencies in the cognitive structure to $n(n - 1)$. Unity then can be defined as:

$$\text{uni} = \frac{\sum_{i=1}^{n} \text{dep}\ (A_i)}{n(n-1)}$$

where $\text{dep}(A_i)$ is the total dependency of the ith attribute.

Degree of organization. Lewin (1951) thought of degree of organization of a whole in terms of the existence of a guiding force or principle that controls its parts. To the extent that one part or a cluster of parts dominates the whole, the whole is said to be highly organized. On the basis of this view the degree of organization can be readily defined. From the dependency matrix the degree to which a given attribute, A_i, depends on others can be computed, as well as the degree to which it determines others. This value is obtained by summing the entries in the appropriate column of the matrix. Taking the determinance of the strongest attribute, $\det(A_i)^{\max}$, and dividing it by unity of the cognitive structure we obtain a measure that reflects the degree to which the interdependence among the attributes is concentrated around a single core. Thus:

$$\text{org} = \frac{\det\ (A_i)^{\max}}{\text{uni}}$$

Empirical measures of the dimensions defined above can be obtained with the aid of various well-known methods that analyze the *S*'s perception or conception of a given stimulus in terms of its characteristics or properties: adjective check lists, rating scales, the semantic differential, *Q* sort, and the like. The method used in the present series of studies required the *S* to describe a stimulus object, namely, another person, by freely listing the qualities and attributes that characterized him. The number of such qualities enumerated constituted the measure of differentiation. Complexity was measured by having the *S* organize the attributes he listed into meaningful groupings and subgroupings, which served as a basis for assigning weights reflecting the respective levels of inclusion. Both unity and organization were computed from *S*s' ratings of dependencies among attributes they listed. Unity was scored by summing the dependency ratings among pairs of attributes, and, as required by the definition, dividing the sum by the maximum possible sum of dependencies. Organization was measured by finding for the given cognitive structure the attribute on which the greatest number of other attributes depended, taking its determinance score (i.e., simply the number of attributes that were rated as depending on it) and dividing it by unity.

The scope of this paper does not allow the discussion of the interrelations among the cognitive dimensions, either as these arise from the formal properties of the definitions or as they were observed in empirical results. An extended analysis of this problem is given elsewhere (Zajonc, 1954).

EXPERIMENT I

Although a person is capable of activating a large number of cognitive structures, at any one time only a few can operate. Specific conditions must therefore exist under which cognitive structures become activated, and the specification of these conditions should enable one to predict the nature of the structures that are activated. It is assumed that a cognitive structure relevant to a particular object or event is activated primarily when the person expects to deal with information about that object or event. Such anticipation constitutes one of the most fundamental occasions for the activation of cognitive structures. Casual observation reveals that messages exchanged between communicators at the onset of communication generally indicate what the communication is going to be about, and thus allow the activation of appropriate cognitive structures. One has merely to scan the opening paragraphs of articles in this journal to observe the tendency. The structure of language also shows such a propensity. Sentences generally begin with the subject. Placing the subject at the beginning of the sentence indicates what cognitive structures should be activated so that the information contained in the predicate can be meaningfully received.

If it is assumed that cognitive structures are activated in the anticipation of dealing with information, it follows that not only the subject matter of anticipated communication but also the person's role in the communication process determines the kind of cognitive structure that he activates. Two basic ways of dealing with information may be distinguished, receiving and transmitting. When a person primarily anticipates receiving information, he may be expected to activate a cognitive structure capable of admitting the incoming information. Concomitant with the anticipation of receiving information is the anticipation of cognitive change. On the other hand, anticipation of transmitting information should activate structures that may serve as a source of potential messages. These two processes may be termed receiving and transmitting tuning. The purpose of this study is to examine the differences in the properties of cognitive structures activated under receiving and transmitting tuning.

Method

Subjects. The *S*s used in this study were recruited on a voluntary basis from the AFROTC at the University of Michigan. Altogether 45 *S*s participated in the experiment.

Procedure. After assembling in a classroom, the *S*s in all conditions were issued the following verbal instructions by *E*:

> This is a study on how groups operate under certain conditions. I will tell you more about it later. Before, however, I will distribute copies of a real letter which was written by one individual to another. I want you to skim over this letter and get a general idea of what sort of person the writer is. Just try to imagine what kind of individual he is, and what are some of the things which are characteristic of him. But, please do not try to memorize the letter. This is not an experiment on memory. Just try to get a general picture of the individual who wrote the letter. You will have two minutes to read it.

At this point the letter was distributed face down; then a signal was given to begin reading the letter. After 2 minutes, a signal was given to turn the letter face down again. The copies of the letter were then collected.

The letter used in the experiment was written to a potential employer by a job applicant. It contained some of the qualifications of the applicant and information regarding his background.

Induction of cognitive tuning. Twenty *S*s received induction intended to produce transmitting tuning, and 25 received induction intended to produce receiving tuning. After the copies of the letter were collected the transmitters received the following instructions:

> I said that we are trying to discover how groups operate under certain conditions. We are especially interested in the process of communication.
>
> At this very moment there is another group in the building. Your responsibility will be to communicate the information you have obtained about the person who wrote the letter to the members of the other group. You will have to describe this person to the other groups, so that they can know him as well as you do now.
>
> But before we begin to transmit this information to the other group, we will put down the things we learned about the writer from his letter. I will now distribute forms which you will fill out according to the instructions written on each page.

The receivers were issued the same instructions except the second and third paragraphs were changed to read:

> At this very moment there is another group in the building. They have detailed information on the individual who wrote the letter, and they will communicate this information to you. We want to see how well they can convey to you all the information they have.
>
> But before we begin to receive this information from the other group, we will put down the things we learned about the writer from his letter. I will now distribute forms which you will fill out according to the instructions written on each page.

TABLE 1

DIFFERENCES IN THE PROPERTIES OF COGNITIVE STRUCTURES BETWEEN TRANSMITTERS AND RECEIVERS

Property	Mean Score for Transmitters	Mean Score for Receivers	t
Differentiation	7.90	5.12	5.67***
Complexity	17.80	9.32	6.83***
Unity	0.309	0.222	2.19*
Organization	13.75	8.40	2.94**

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Measurement of the properties of cognitive structures. To obtain measures of the properties of cognitive structures the following procedure was used. *S*s were issued instruction booklets to which were attached stacks of 52 blank cards marked from A to ZZ. The *S*s' first task was outlined on the first page of the booklets:

> Detach the stack of cards from this booklet.
>
> On each card *separately* write *one* characteristic which describes the applicant. You can put down whatever comes to your mind, since there is no one list of characteristics that can be considered as either "correct" or "incorrect." Everyone of us sees things in a slightly different way.
>
> You may have too many or too few cards, but this shouldn't bother you. Put down as many characteristics as you feel are necessary to describe the applicant adequately. Work rapidly.

The number of characteristics attributed to the applicant constituted the degree of differentiation in each case.

To determine complexity the *S*s were required to:

> Lay out in front of you all the cards you used for listing the characteristics of the applicant. Look over them carefully and notice whether they fall into some broad natural groupings. If they do, arrange them into such groups.

After arranging the cards into groupings, the *S*s were issued the following instructions:

> Now, look at your groups one by one and see whether these can't be broken down into subgroups. If they can, separate the cards accordingly. It is also possible that these subgroups can be broken down further, and so on.
>
> When you arrange *all* cards into groups and subgroups, list your groupings on the sheet below as if they were points and subpoints of an outline. First, give names or titles to your groups and subgroups. Then in the right-hand column list the letters of all the characteristics that belong into the respective groups and subgroups.

The groupings and subgroupings were used to determine the levels of inclusion of the attributes, the basis for computing complexity scores.

The *S*s were instructed to proceed as follows for the last task:

> It is possible that some of the characteristics are related to one another. They may depend on one another in such a way that if one changes, the other ones would change too. Suppose the table in front of you were bigger than it is now. Then it would also become heavier. This means that the weight of the table depends on its size. The relationships between the characteristics you put down may not be so obvious and so simple, but try to decide whether such relationships exist nevertheless. To do this lay out the cards in front of you in alphabetical order, and follow the procedure below:
>
> List all the characteristics which would change if Characteristic *A* were changed, absent, or untrue of the applicant; list all the characteristics which would change if Characteristic *B* were changed, absent, or untrue of the applicant, etc.

On the basis of the above responses dependency matrices were constructed and measures of unity and organization computed.

After the *S*s completed the booklets they were told with an apology that the experiment was over.

Results and Discussion

Degree of differentiation. The overall results are shown in Table 1, which shows a significantly higher degree of differentiation for transmitters. The transmitters listed, on the average, 7.90 attributes, the receivers 5.12. It was assumed that the components of cognitive structures activated under tuning for reception act as *receiving categories*, whereas those activated under tuning for transmission act as *potential units of transmittable information*. The relevant difference between these two types of attributes is along the dimension of specificity–generality. Attributes whose function is to admit all possible types of information must be sufficiently broad and general. Since they are to serve as "codes" or "files" in terms of which ever new units of information are admitted, they must be sufficiently comprehensive so that all possible information can be appropriately categorized.

When the components of the cognitive structure are to serve as potential units of transmittable information they are more likely to be specific. A "message" is composed of specific units alone, whereas general units constitute skeletons that organize the substantive and specific units of the message. Since specific units require more detail, their essential difference from general units lies in the coarseness of their "grain." Cognitive representations made in specific terms normally require a larger number of attributes than representations made primarily in broad and general terms. If it is assumed that, other things equal, differentiation is a function of the level of specificity of a given cognitive structure, and if tuning for transmission results in a tendency toward greater specificity than tuning for reception, transmitters should indeed show higher differentiation. Thus, if it can be demonstrated that the transmitters in fact employ a higher proportion of specific attributes than the receivers, then the differences in the degree of differentiation are better understood.

The attributes of all *S*s were coded for specificity by two independent judges, assigning each attribute to either specific or general categories. The general category included such items as "intelligent," "shiftless," "immature," "he does not seem to be ambitious." The specific category included attributes like, "didn't finish college because of financial problems," "he has never worked as a mail clerk." The reliability measured in percent agreement was 82%. The results indicate that the transmitters' cognitive structures had on the average 57.6% specific attributes, while those of the receivers, only 32.6%. This difference was significant at the .001 level.

Degree of complexity. The results show a higher degree of complexity for transmitters. Two factors may account for this finding. First, it would appear that the larger proportion of specific attributes, the more likely should the attributes fall into equivalence classes and subclasses. Furthermore, the results have indicated a higher degree of differentiation on the part of the transmitters. From the definition of complexity it is evident that a high degree of differentiation increases the probability of high complexity. Both factors work to enhance complexity on the part of the transmitters.

Degree of unity. From the concept of cognitive tuning it would follow that receiving tuning corresponds to the anticipation of a cognitive change in one's own cognitive structure, whereas transmitting tuning corresponds to the anticipation of inducing changes in cognitive structures of other persons. A person who anticipates the occurrence of cognitive change is very likely to "unfreeze" his cognitive structure or to make it more susceptible to the anticipated changes. Susceptibility and resistance to change may be thought of in terms of the strength of forces necessary to induce changes in a given cognitive structure. In order to change a highly unified structure a strong force is necessary because the components of the structure depend on one another to a great degree. Thus a change in one attribute should result in changes in other attributes. In order to change a single attribute of a highly unified cognitive structure, a force must be applied which is capable of overcoming the resistance of not only the given attribute, but of all the other attributes on which it depends. To produce a similar change in an un-unified structure, a force is required that is capable of overcoming the resistance of the given component alone. If cognitive structures activated under receiving tuning are more susceptible to change than those activated under transmitting tuning, then they must be less well unified than the latter. The results indicate that this, in fact, is the case.

Degree of organization. The dynamics of receiving tuning are of such nature that a high degree of structure and elaboration should deter the smooth reception of information. Rather, the situation calls for unstructured and loosely integrated cognitive structures, so that all possible types of changes could easily be introduced. The receiver does not know what specific information is to be imparted to him, and he cannot focus his cognitive processes on any single aspect of the matter, but must think in general terms. Degree of organization, it will be recalled, was defined in terms of the extent to which the cognitive structure is organized around such one specific aspect. Receiving tuning should therefore result in a relatively low degree of organization.

Transmitting tuning, however, represents a different psychological state. The transmitter is aware of the specific units of information to be transmitted, and his thinking revolves around them, with a certain focus of attention. He should therefore show a relatively higher degree of organization.

One should also expect a higher degree of organization for transmitters on other grounds. Predominance of specific attributes requires a structure into which the attributes can be ordered. The specific attributes are organized by means of general ones. Since the transmitters were shown to employ fewer general attributes, their cognitive structure should be organized around fewer foci than those of the receivers.

The results show that the degree of organization was significantly higher for transmitters than for receivers. It should be noted that this result was obtained independently of the differences in unity. From the definition it follows that, other things being equal, degree of organization actually decreases with higher unity. The higher degree of organization found for the transmitters was obtained despite their higher unity.

EXPERIMENT II

In the previous experiment lower degrees of differentiation, complexity, unity, and organization were obtained under receiving than under transmitting tuning. These results were assumed to derive from the fact that the receivers tune in flexible cognitive structures, readily susceptible to changes, and having few specific attributes, while the transmitters' cognitive structures are characterized by a considerable resistance to change and a larger proportion of specific attributes. If resistance to change and degree of specificity account for the differences arising under receiving and transmitting tuning, they should also account for data obtained under other conditions involving these two factors. Suppose that persons expect to enter into communication with others who hold opinions opposite to or incongruent with their own. In the case of receiving tuning, there should be more resistance to accepting information that is contrary to the person's own beliefs than to accepting information that supports them. Since resistance to cognitive change has been assumed to be reflected in increased unity, we expect the receivers to show a relatively higher degree of unity when dealing with incongruent information than when dealing with congruent information. Furthermore, if they are to protect themselves from unwanted changes, receivers of incongruent information should eliminate from their cognitive structures components that make them vulnerable to change. If they have committed themselves to a positive opinion, they should ignore or omit negative attributes, and if they have committed themselves to a negative opinion, the opposite should be the case. Jones and Kohler (1958) have recently shown that *S*s tend to recall plausible arguments that support their opinion and implausible arguments that question it, and to forget the converse. Simultaneously, the commitment to an opinion places a requirement on specificity. Thus the first factor, because of its *selective* effects should tend to reduce differentiation for receivers of incongruent information, while the second should tend to increase it. If the two factors are equal in magnitude, the degree of differentiation should remain the same for receivers regardless of whether they deal with congruent or incongruent information. Similar consequences should follow for transmitters. The differences in the degree of differentiation between receivers and transmitters when incongruent information is involved should depend therefore on the relative increments of specificity and selectivity in both groups. If we assume that both groups select all those and only those attributes which are consistent with their position, then because the number of such attributes is held constant for both groups, no differences in specificity should appear, and, consequently, the groups should not manifest any differences in differentiation. If receivers and transmitters increase in specificity and selectivity when confronted with incongruent information, with these increments being more pronounced for receivers, we should expect a lowering of differentiation for transmitters and a heightening for receivers. The differences in complexity should show the same pattern as differences in differentiation. Degree of organization, on the other hand, should increase for both groups, since now there is a core around which cognitive components may readily become organized—namely, the commitment to a positive or a negative position. But again this increase should be relatively more pronounced for the receivers, resulting in approximately equal levels of organization for both groups. The purpose of the second experiment is to examine these predictions.

Method

Subjects. The *S*s were sophomore and junior women enrolled at the University of Michigan. Altogether, 92 *S*s took part in the experiment.

Procedure. Four groups of 23 *S*s each were assigned randomly to four experimental groups: Transmitters of Congruent Information (T), Receivers of Congruent Information (R), Transmitters of Incongruent Information (TI), and Receivers of Incongruent Information (RI). The stimulus materials and the measurement method were the same as in Experiment I. Also, the procedure for the T and R groups was the same as in Experiment I, except that for the present purposes the induction of tuning preceded the reading of the letter. The TI and RI groups were issued the following instructions after assembling in the experimental room:

> The main point of this study is to find out something about the process of communication. I will give you a letter to read. This letter was written by an individual who is applying for a job. After you have read this letter and have gotten some idea about its author, each one of you will decide for herself whether or not he should be hired. Let me tell you something else. Right now there is another group in the building. They have the same information about the applicant, and each of them is also making a decision whether or not the individual should be hired. After you have decided whether or not to hire the applicant, every one of you will be paired with a member of the other group. What is more, you will be paired up with someone who decided differently from you. In other words, if a member of this group had decided to hire him, she will be paired with someone who has decided not to hire him, and vice versa.

Both the TI and the RI groups were given these initial instructions, and, afterwards, transmitting and receiving tuning was induced in the same manner as it was in Experiment I. The procedure for collecting data was the same as in the previous experiment.

Results

It was assumed that anticipation of dealing with incongruent information and commitment to an opinion result in an increased proportion of specific attributes, especially for the receivers. The average percentages of specific attributes for the T, R, TI, and RI groups, respectively, were 52.0%, 39.8%, 72.3%, and 70.7%. The analysis of variance (using the arc sin $\sqrt{\%}$ transformation) presented in Table 2 shows significant effects for both tuning and congruence. It will be noted that under conditions that involve anticipation of dealing with incongruent information, both receivers and transmitters increase the proportion of specific attributes markedly, with the increase being somewhat larger for the receivers. It is also clear from the data that when persons anticipate dealing with contrary information, the differences in the proportion of specific attributes between the groups disappear. The difference in the proportion of specific components between the R and the T groups was significant at the .01 level, and not significant between the RI and TI groups.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE FOR THE PROPORTION OF SPECIFIC ATTRIBUTES

Source	*df*	*MS*	*F*
Tuning (T)	1	3,885.61	5.85*
Congruence (C)	1	8,068.15	12.14***
T × C	1	1,585.74	2.39
Error	88	664.27	

* $p < .05$.
*** $p < .001$.

Degree of differentiation. The increment in specificity alone would lead to the expectation that the degree of differentiation should be greater for both groups that dealt with incongruent information. The receivers of contrary information should show a more pronounced increase, such that they reach the level of differentiation equal to that of the TI group. However, it was assumed that a factor operates simultaneously that tends to reduce differentiation, namely, the tendency to select attributes consistent with one's position. If this factor operates equally for both groups, as was assumed above, the TI groups should show a decrement and the RI group an increment. The results for all the properties, as summarized in Table 3, follow just this pattern. The analysis of variance, presented in Table 4, shows significant main effects deriving from tuning as well as a significant T × C interaction. The overall mean differentiation scores are somewhat higher than when *S*s deal with congruent information, but this difference is not significant. The differences between transmitters and receivers are significant only when congruent information is involved ($p <$

TABLE 3

DIFFERENCES IN THE PROPERTIES OF COGNITIVE STRUCTURES BETWEEN TRANSMITTERS AND RECEIVERS DEALING WITH CONGRUENT AND INCONGRUENT INFORMATION

Property	Transmitters		Receivers	
	Congruent	Incongruent	Congruent	Incongruent
Differentiation	11.39	9.26	6.65	7.70
Complexity	27.09	20.83	14.35	16.48
Unity	0.265	0.269	0.226	0.300
Organization	15.54	22.13	10.41	19.48

TABLE 4

SUMMARY OF ANALYSES OF VARIANCE FOR DATA IN TABLE 3

Source	*df*	Differentiation		Complexity		Unity		Organization	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Tuning (T)	1	228.53	30.92***	1678.79	21.15***	.027	1.35	568.13	7.12**
Congruence (C)	1	6.79	<1.00	198.10	2.50	.035	1.75	1034.51	12.97***
T × C	1	64.72	9.13***	304.88	3.48*	.001	<1.00	151.33	1.90
Error	88	7.39		79.36		.020		79.75	

* $p < .05$ level.
** $p < .01$ level.
*** $p < .001$ level.

.001). It appears, therefore, that there must exist strong selective effects, as well as increments in specificity when persons deal with contrary information, and that these two factors result in the equalization of differentiation.

Degree of complexity. The results for complexity are essentially the same as for degree of differentiation. The effects of tuning on complexity were significant ($p < .001$), as was the T × C interaction ($p < .05$). The degree of complexity of the transmitters shows a drop when they anticipate dealing with contrary information. The difference between the T and the TI group is significant at the .01 level. The decrement in complexity on the part of the transmitters is due to the selective effects of commitment to a position. Since they abandon those aspects that may weaken their position, their cognitive structures lose in elaboration. The receivers show a slight but insignificant increment in complexity, and a significant T × C interaction is thus obtained. It should be noted, however, that the TI group shows a higher degree of complexity than the RI group, this difference being significant at the .05 level. Apparently, the effects of tuning are sufficiently strong here to be maintained despite the selectivity and the specificity factors.

Degree of unity. While there is some increase in the degree of unity in the "incongruent" condition, the overall effects of congruence do not reach an acceptable level of significance. The strongest difference appears to be that between the R and the RI group. This was the only significant result with respect to unity ($p < .05$). It may be assumed, therefore, that when dealing with incongruent information receivers protect themselves from unwanted changes by increasing unity, while transmitters because of their already high unity need not increase it.

Degree of organization. As expected, degree of organization shows pronounced effects of both tuning and congruence. For both, receivers and transmitters, there are significant increments in organization when contrary information is involved (both differences are significant at the .01 level). The commitment to a position, therefore, because it leads to the selection of a particular type of material—selection that is quite systematic—provides a strong core around which the components of the cognitive structure may become readily organized. Even though the receivers showed an increase in unity when dealing with contrary information, they manifest a significant, almost twofold, increment in degree of organization. The effects of the anticipation of dealing with incongruent information are therefore stronger for the receivers than for transmitters.

SUMMARY

The cognitive effects of the person's role in the communication process were examined in two experiments. A systematic method for the description of cognitive structures was developed, and in the first experiment persons expecting to transmit information were compared with others expecting to receive information for the extent of differentiation, complexity, unity, and organization. The results show that transmitters activate cognitive structures which are more differentiated, complex, unified, and organized than those activated by receivers.

The second experiment involved the anticipation of dealing with incongruent information, and in comparison with groups dealing with congruent information, those expecting to deal with incongruent information generally showed decreased differences between transmitters and receivers. These results were accounted for by an increase in the proportion of specific cognitive components and in the tendency to reject material inconsistent with the person's own opinion.

REFERENCES

ACH, N. *Ueber die Willenstätigkeit und das Denken.* Gottingen: Yandenhoeck & Ruprecht, 1905.

ASCH, S. E. Forming impressions of the personality. *J. abnorm. soc. Psychol.*, 1946, **41,** 258–290.

BRUNER, J. S., GOODNOW, J., & AUSTIN, G. *A study of thinking.* New York: Wiley, 1956.

CHAPMAN, D. W. Relative effects of determinate and indeterminate Aufgaben. *Amer. J. Psychol.*, 1932, **44,** 163–174.

FRENCH, V. The structure of sentiments. *J. Pers.*, 1947, **15,** 247–280.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion: Psychological studies of opinion change.* New Haven: Yale Univer. Press, 1953.

JONES, E. E., & KOHLER, R. The effects of plausibility on the learning of controversial statements. *J. abnorm. soc. Psychol.*, 1958, **57,** 315–320.

KULPE, O. Versüche über Abstraction. *Ber. I. Kongr. Exp. Psychol.*, 1904, 56–68.

LEWIN, K. Das Problem der Willenmessung und das Grundgesetz der Assoziation. *Psychol. Forsch.*, 1922, **1,** 191–302. (a)

LEWIN, K. Das Problem der Willenmessung und das Grundgesetz der Assoziation. *Psychol. Forsch.*, 1922, **2,** 65–140. (b)

LEWIN, K. *Field theory in social science.* New York: Harper, 1951.

WATT, H. J. Experimentelle Beiträge zur einer Theorie des Denkens. *Arch. ges. Psychol.*, 1905, **4,** 289–436.

ZAJONC, R. B. Cognitive structure and cognitive tuning. Unpublished doctoral dissertation, University of Michigan, 1954.

ZIMMERMAN, CLAIRE, & BAUER, R. A. The effect of an audience upon what is remembered. *Publ. opin. Quart.*, 1956, **20,** 238–248.

(Received August 10, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 168-175

THE ROLE OF UNCERTAINTY IN COGNITIVE CHANGE[1]

ROBERT B. ZAJONC AND JULIAN MORRISSETTE[2]

University of Michigan

IN analyzing the process of social change in small groups Lewin (1952) proposed that "a successful change includes . . . three aspects: unfreezing (if necessary) the present level L_1, moving to the new level L_2, and freezing group life on the new level." He argued that since stable states resist changes the stability of the group must be reduced before a permanent change can be introduced. These components of the process of changing appear to have a rather general application. Individual as well as social change might be so characterized. The processes of attitudinal and cognitive change, for instance, may well involve the three steps distinguished by Lewin. The literature on the Water-Jar Einstellung Test (Levitt, 1956) shows some evidence that the ability to see the new solution is related to variability on previous solutions. That is, the likelihood of a new strategy depends on the rigidity of the old strategy.

The purpose of this study is to determine whether cognitive change involves similar dynamics. The specific cognitive variable to be examined is cognitive uncertainty. It has been shown that cognitive change and cognitive uncertainty are intimately related, and in particular, that the likelihood of cognitive change increases with uncertainty (Zajonc & Burnstein, 1957). It is not clear, however, whether cognitive change occurs *only* with a concomitant increase in uncertainty. According to Lewin's conception of the dynamics of change, it is hypothesized that uncertainty must increase before a cognitive change can take place. Of course, when the person is very uncertain to begin with, no such increase is to be expected, and we should therefore expect that increases in uncertainty accompanying cognitive change are more likely to occur at low levels of uncertainty. But given low uncertainty, cognitive change can be accomplished only by previously raising uncertainty.

METHOD

Subjects

Ninety-six male *S*s recruited on a volunteer basis from the undergraduate classes at the University of Michigan participated in the experiment. Eight groups of 12 *S*s each were employed.

Procedure

In all groups the *S*s' task was to estimate the number of bomb craters shown in a slide of an aerial photograph. In each group, the slide was presented 18 times tachistoscopically for intervals of 0.5 sec.

Three methods of inducing cognitive change were employed: No Feedback (NF), Feedback I (FI), and Feedback II (FII). In the NF condition *S*s estimated the numbers of bomb craters on each of the 18 presentations and no feedback as to their accuracy was ever given. Three experimental groups constituted the NF condition, NF_a, NF_b, and NF_c. In the NF_a group the same slide was presented in all of the 18 trials. In the NF_b group the experimenter (*E*), without the *S*s' awareness, changed the slide following the sixth trial, and in the NF_c following the twelfth trial. The original slide pictured 350 craters, the new slide 250. Both slides showed the same area, but in the second slide 100 bomb craters were eliminated by retouching.

In the FI groups the slide containing 350 craters was presented throughout all 18 trials. After each trial the *E* would announce a number representing the mean estimate of a group of pilots who supposedly were previous *S*s of the experiment. Again three experimental groups were employed: FI_a, FI_b, and FI_c. In the FI_a group the estimate reported by the *E* was always around 350. The mean of the estimates given by the *E* over the 18 trials was 350. On any trial the deviation from 350 was never greater than 10. In the FI_b group the reported mean changed from 350 to 250 following the sixth trial, and in the FI_c following the twelfth trial.

In the FII condition the feedback was held constant at around 350 craters for all 18 trials. In these groups, however, the slide was changed as in the NF condition. Since the FI_a group provides a baseline for the FII groups, only two experimental groups were employed, FII_b and FII_c. In the FII_b the slide changed following the sixth trial, and in the FII_c following the twelfth trial.

In all groups the *S*s were required to write down their judgments as to the number of bomb craters presented and also to indicate the smallest range within which their estimates would be perfectly accurate. This procedure provided a subjective measure of uncertainty (SU) which reflects the *S*s' own estimates of their errors. Another measure of uncertainty was obtained by employing the variability of *S*s' estimates as to the number of bomb craters. The trials were grouped into six trial blocks consisting of three successive trials each, and for each *S*, average deviations of the three estimates within each trial block were computed. This latter, more objective, measure of uncertainty will be referred to as the OU measure.

---

[1] These studies were a part of a larger program supported by Grant G-2802 from the National Science Foundation.

[2] Now at Wright-Patterson Air Force Base.

RESULTS AND DISCUSSION

Decrement in Uncertainty (OU Measure) as a Function of Repeated Trials

Before considering the role of uncertainty in cognitive change the data from the two control groups, NF$_a$ and FI$_a$ will be examined. The results for the two groups are presented in Figure 1 and the analysis of the data in Table 1.

In both F and NF conditions, there was a pronounced decrement in uncertainty as a function of repeated exposures of the stimulus. The analysis of variance shows an *F* ratio of 15.00, significant at the .001 level, for differences between Trial Blocks. It is also clear from Figure 1 that while the FI$_a$ group shows a considerably higher level of uncertainty at the outset of the session, its uncertainty falls below that of the NF$_a$ group beginning with the second trial block. The *t* test between the two groups on Trial Block I was significant at the .01 level. Thus we should expect a significant interaction between the effects of Feedback and Trial Blocks, as the analysis of variance in fact confirms.

TABLE 1

ANALYSIS OF VARIANCE FOR DATA IN FIGURE 1

Source	*df*	*MS*	*F*
Feedback	1	881	<1.00
*S*s in same group	22	1,002	
Trial blocks	5	12,010	15.00***
Trial blocks × Feedback	5	2,298	2.87*
Pooled *S*s × Trial blocks	110	800	

* $p < .05$.
*** $p < .001$.

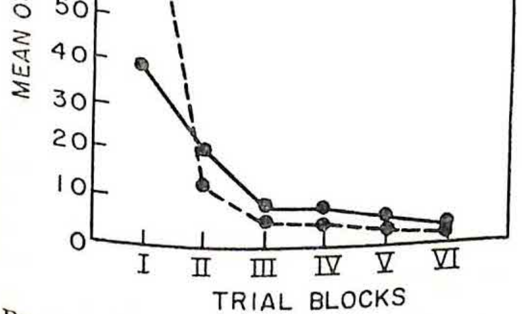

FIG. 1. Changes in uncertainty (OU Measure) as a function of repeated exposures of the stimulus.

TABLE 2

ANALYSIS OF VARIANCE FOR DATA IN FIGURE 2

Source	*df*	*MS*	*F*
Feedback	1	158	<1.00
*S*s in same group	22	52,234	
Trial blocks	5	20,778	7.83***
Trial blocks × Feedback	5	8,909	3.36**
Pooled *S*s × Trial blocks	110	2,654	

** $p < .01$.
*** $p < .001$.

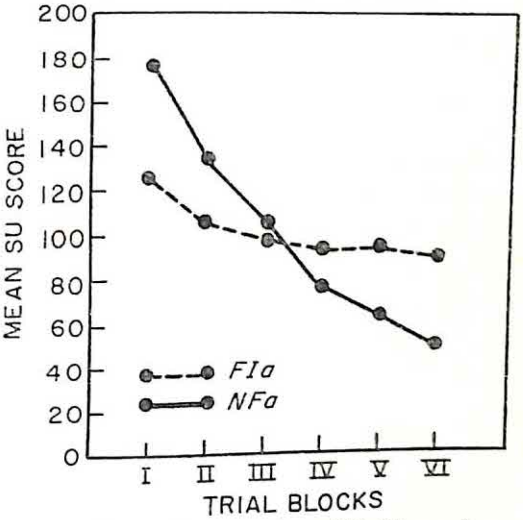

FIG. 2. Changes in uncertainty (SU Measure) as a function of repeated exposures of the stimulus.

Decrement in Uncertainty (SU Measure) as a Function of Repeated Trials

When uncertainty is measured by the *S*s' estimates of their error, the decrement over Trial Blocks is again observed, but now the interaction between Feedback and Trial Blocks is reversed. It is clear from Figure 2 that the FI$_a$ curve has a lesser slope than the NF$_a$ curve, and that at the outset uncertainty of the NF$_a$ group is higher than that of the FI$_a$ group. The analysis of variance presented in Table 2 indicates again a significant effect deriving from Trial Blocks as well as a significant interaction between Trial Blocks and Feedback.

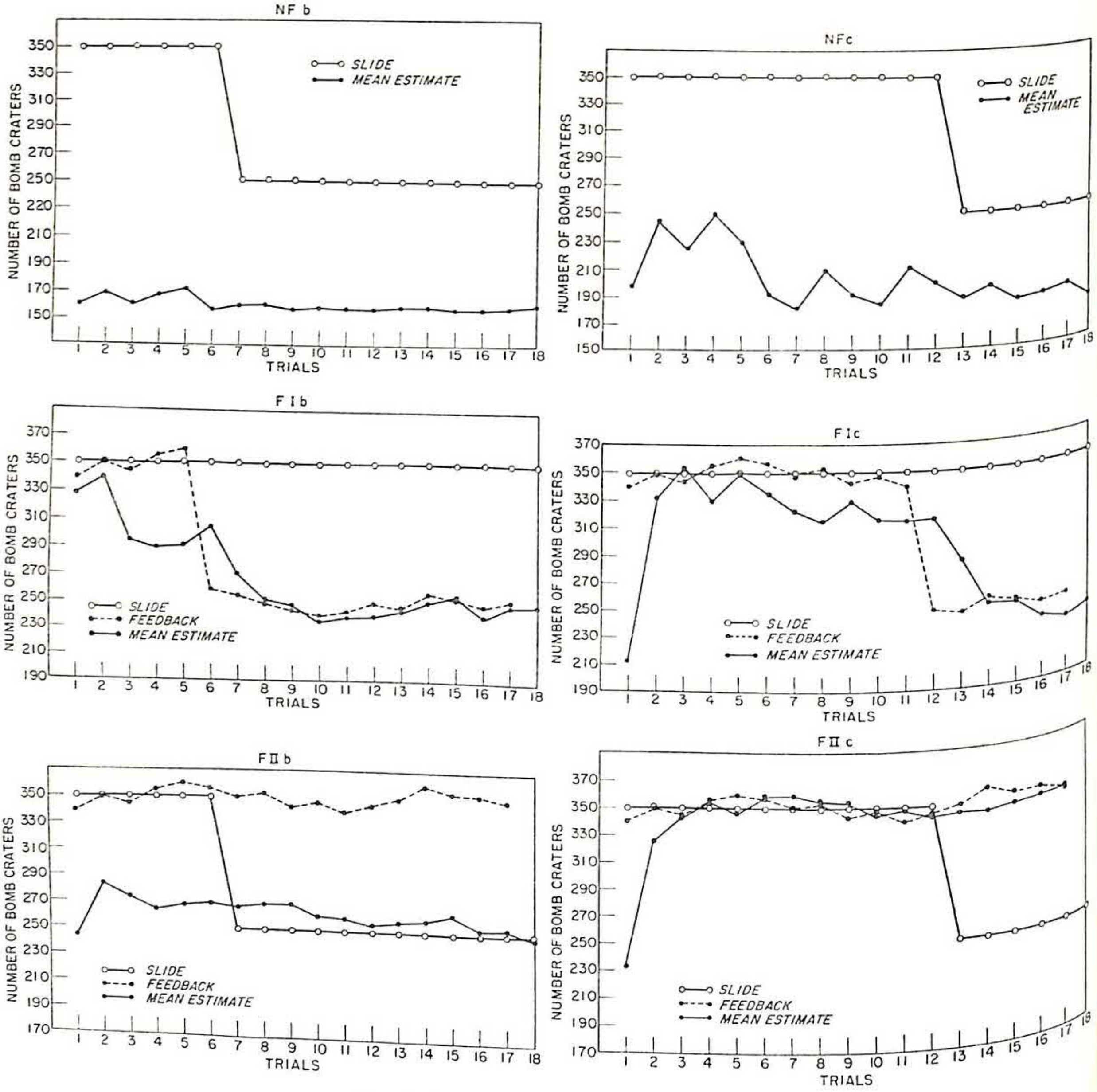

FIG. 3. Mean estimates of bomb craters.

The difference in the direction of interaction is probably a function of the uncertainty measures employed. Since the OU measure is essentially a measure of intertrial and intrasubject variability we should expect that feedback would originally produce high judgmental variability and thus high OU scores. The first feedback occurs after the first trial of the first trial block, after *S* has already made his judgment. If feedback has any effect on *S*'s judgments it should result in a change of his judgment on the immediately subsequent trials, and therefore a high average deviation on the first trial block. Moreover, since the feedback maintains the judgments around a certain level, the remaining trials should exhibit low variability. In the NF_a condition there are no comparable forces on the *S*s to change their estimates during the first three trials. Nor is there pressure toward the maintenance of the judgments throughout the remaining trials. The NF_a condition therefore exhibits an initially low level of OU scores and, although uncertainty decreases over Trial Blocks, this decrement is less abrupt than in the FI_a group.

The SU measure, however, reflects the *S*s' judgmental confidence. If feedback serves as a source of additional information it should increase confidence and result in the FI_a group in relatively low SU scores on the initial trials. Thus, during the first three trial blocks the SU scores are lower for the FI_a group than for the NF_a group. However, the FI_a group does not decrease as rapidly as the NF_a group, a result that is difficult to interpret. It implies that acceptance of the additional information provided by the feedback may reduce uncertainty at the outset, and simultaneously acts to maintain the resulting level of uncertainty through the remainder of the session.

Uncertainty and Cognitive Change

The data thus far presented show that repeated exposures to a stimulus result in uncertainty that decreases progressively. Both the person's judgmental variability and his subjective estimate of error decrease significantly over repeated trials. The "unfreezing" hypothesis suggests that if the individual faces a change of judgment, this change is concomitant with increased uncertainty. Moreover, if we assume that there exists a level of uncertainty at which cognitive change can readily occur, and below which cognitive change requires an increase in uncertainty, it follows that a cognitive change on the third trial block should be accompanied by a lesser increase in uncertainty than on the fifth trial block. However, before introducing results on the role of

TABLE 3

MEAN ESTIMATES OF BOMB CRATERS BEFORE AND AFTER STIMULUS AND FEEDBACK CHANGE

Condition	Mean Estimate of Bomb Craters		Difference
	Before	After	
NF_b	171.42	159.67	−11.75
NF_c	209.67	184.08	−25.59
FI_b	308.33	247.83	−60.50
FI_c	321.92	251.17	−70.75
FII_b	271.08	259.92	−11.16
FII_c	339.42	346.17	6.75

TABLE 4

ANALYSIS OF VARIANCE FOR DATA IN TABLE 3

Source	*df*	*MS*	*F*
Feedback (F)	2	206,499	24.03***
Time of change (T)	1	4,834	<1.00
T × F	2	14,747	1.72
Ss in same group	66	8,594	
Trial blocks (B)	1	29,929	9.01**
B × F	2	12,995	3.91**
B × T	1	38	<1.00
B × F × T	2	909	<1.00
Pooled *Ss* × B	66	3,322	

** $p < .01$.
*** $p < .001$.

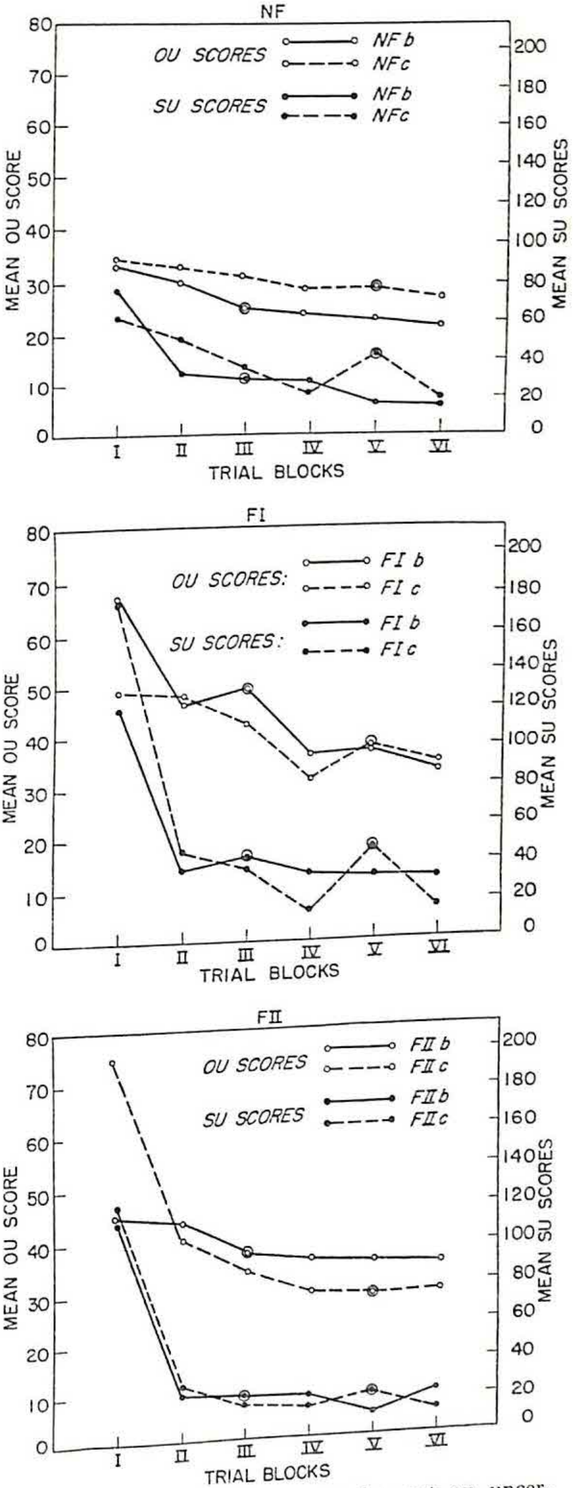

FIG. 4. Effects of change requirement on uncertainty. (Points circled indicate where the requirement for change was introduced.)

TABLE 5

ANALYSIS OF VARIANCE FOR DATA IN FIGURE 4

Source	OU Measure			SU Measure	
	df	*MS*	*F*	*MS*	*F*
Overall Trend	5	13,612	35.59***	34,874	16.03***
Between Group Means	(5)	(857)	(1.23)	21,717	.44
Feedback	2	1,930	2.76	49,315	1.01
Time of change	1	425	.61	137	.00
Interaction	2	1	.00	4,913	.10
Between Group Trends	(25)	(615)	(1.61)*	(3,406)	(1.57)*
Linear	(5)	(1,057)	(1.26)	(8,040)	1.42
Feedback	2	1,733	2.06	7,893	1.39
Time of change	1	798	.95	583	.10
Interaction	2	511	.61	11,916	2.10
Quadratic	(5)	(1,023)	(1.76)	(4,674)	(1.77)
Feedback	2	2,062	3.54*	4,774	1.81
Time of change	1	209	.36	1,652	.62
Interaction	2	392	.67	6,086	2.30
Cubic	(5)	(502)	(1.67)	(1,884)	(1.09)
Feedback	2	921	3.06	1,066	.62
Time of change	1	248	.83	165	.09
Interaction	2	209	.70	3,561	2.07
Quartic	(5)	(331)	(2.31)	(1,433)	(1.89)
Feedback	2	405	2.83	522	.69
Time of change	1	678	4.74*	419	.55
Interaction	2	83	.58	2,851	3.77*
Quintic	(5)	(162)	(3.49)**	(999)	(7.51)***
Feedback	2	464	9.98***	792	5.95**
Time of change	1	228	4.91**	936	7.03**
Interaction	2	27	.59	1,237	9.30***
Between Indiv. Means	66	699	1.83**	49,006	22.52***
Between Indiv. Trends	(330)	(383)		(2,176)	
Linear	66	841		5,667	
Quadratic	66	582		2,644	
Cubic	66	301		1,724	
Quartic	66	143		757	
Quintic	66	46		133	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

uncertainty in cognitive change we shall consider the effects of the experimental induction intended to produce cognitive change.

The effects of stimulus change and feedback change on judgments. The groups in which an attempt was made to introduce cognitive change were NF_b, NF_c; FI_b, FI_c; FII_b, FII_c. The b groups received induction to change on the third trial block and the c groups on the fifth. In the NF condition the experimental induction of cognitive change was introduced by changing the slide. In the FI groups the slide remained the same throughout the session but feedback changed. And in the FII groups the slide changed while feedback remained constant. The *S*s' average estimates of the number of bomb craters for the individual trials are presented in Figure 3. It is clear from the data that the only groups where change of any magnitude occurred as a result of the induction were the FI_b and FI_c groups. It is also evident that in all but the FII_b groups there was a general tendency to conform to the information provided in feedback rather than to that given in the slide.

The results pertaining to cognitive change were evaluated by comparing estimates for the two trial blocks in which change induction was introduced. The average estimates for the trial block immediately preceding and the trial block immediately succeeding the change induction are shown in Table 3 and the analysis of variance on these results in Table 4. The data of main interest are of course the mean estimates for the critical trial blocks. It is clear that there was some overall change in the direction of the slide and feedback changes. The differences between the critical trial blocks are significant at better than the .01 level. The strongest difference, however, occurred in the FI condition and the weakest in the FII condition. The interaction between Trial Blocks and Feedback is significant at the .05 level. These results suggest that the effects stemming from Feedback were more pronounced than those from the stimulus itself. Since the *S*s viewed the slide under suboptimal conditions, they would, of course, tend to rely more on information provided in Feedback. The analysis also shows that in the absence of Feedback *S*s generally underestimated the number of bomb craters in the aerial photograph, and consequently the overall effects of Feedback are significant. The difference in estimates between Trial Blocks IV and V is somewhat greater than between Trial Blocks II and III, but the interaction between Time of Change and Trial Blocks was not significant.

Changes in uncertainty as a function of stimulus change and feedback change. The results for judgments showed that only the FI groups underwent a cognitive change of any significance. The NF and the FII groups changed their estimates only slightly with the FII condition showing the weakest change. We would therefore expect that increments in uncer-

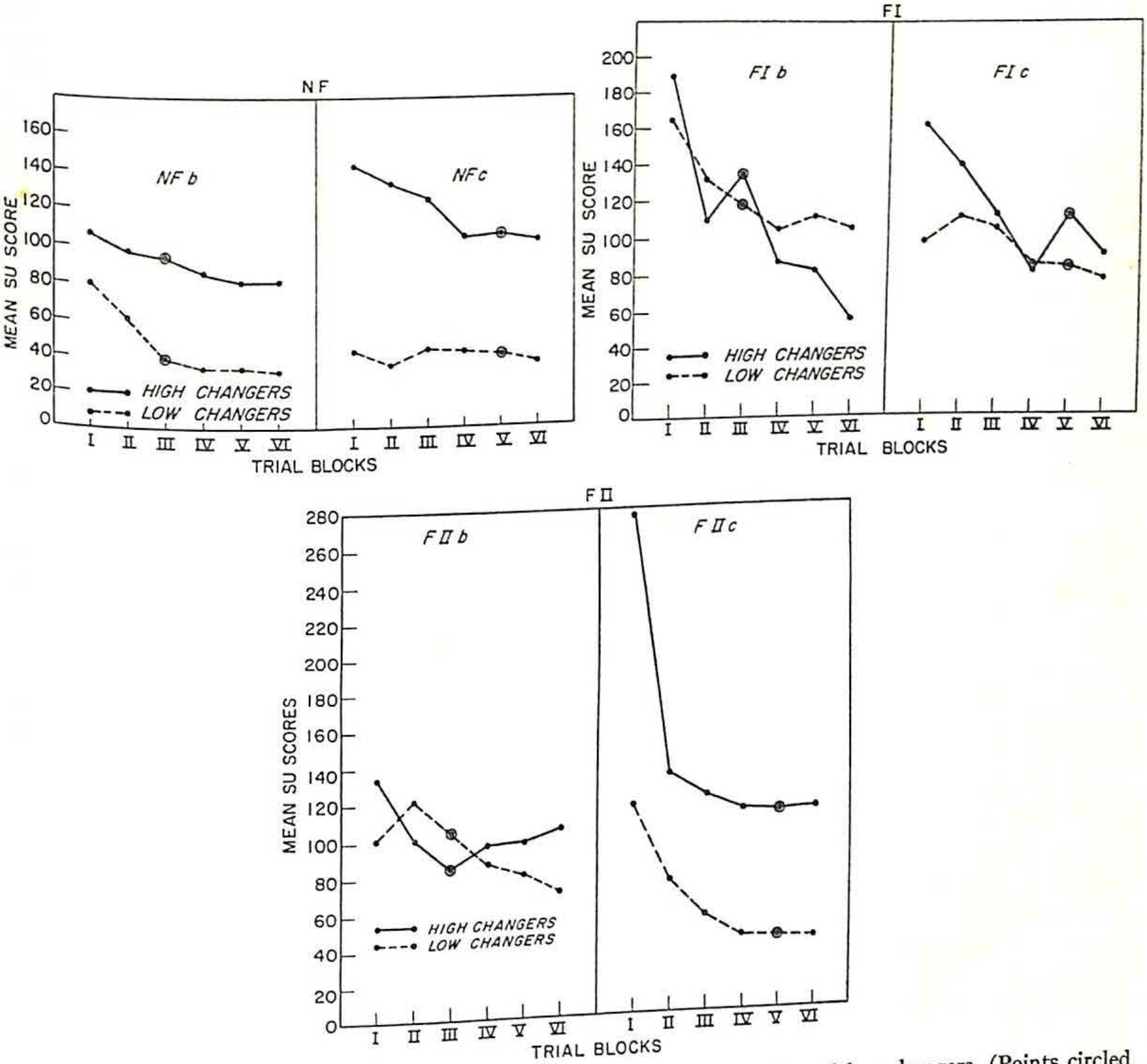

FIG. 5. Effects of change requirement on uncertainty (SU scores) for high and low changers. (Points circled indicate where the requirement for change was introduced.)

tainty at the point of change would occur in corresponding order. The data for all groups in which the attempt was made to introduce cognitive change are shown in Figure 4. Since the hypothesis predicts a departure from monotonicity, orthogonal polynomial trend analyses (Grant, 1956) were performed on the data and the results are shown in Table 5.

The findings are essentially the same for the OU and the SU scores. The curves show that when change in the slide or in feedback was introduced early during the session, a rise in uncertainty occurred only in the FI_b group. However, when the requirement for change occurred during the fifth trial block, all conditions showed an increment in the OU scores, and all but the FII_c group increments in SU scores. Since the level of uncertainty is higher at Trial Block II than at Trial Block V, cognitive change may occur then without much increment in uncertainty. However, at the close of the session when the level of uncertainty is at a minimum and the judgments have become stabilized, cognitive change requires an increase in uncertainty.

The trend analysis shows that the effects of Feedback and the Time of Change are significant for the quintic component. This finding is consistent with the prediction since the summated curves for the b and c groups generate a function with three maxima. One maximum obtains at Trial Block I, the other at Trial Block III where the early change is introduced, and the third at Trial Block V, where the late change is introduced. It should be noted that strong individual differences are evident in the analysis, especially in the SU measure.

Changes in uncertainty as a function of cognitive change. The analysis of variance showed significant effects deriving from individual differences. A close examination of individual results disclosed that there were both considerable differences between the overall uncertainty levels of different individuals and differences in the uncertainty increments. Moreover, some individuals exhibited pronounced changes in judgments while the judgments of others remained stable throughout the entire session. Did the increments in uncertainty occur only for individuals who changed their estimates, or were uncertainty increments distributed equally over all *S*s? Changes in the direction of the induction were computed for each *S* and *S*s in each condition were divided at the median into high and low changers. Since the most pronounced individual differences were revealed by the SU measure, analysis was performed using the SU scores. Figure 5 shows mean SU scores for the six trial blocks for the high and low changers.

It is clear from the curves that *S*s who changed their estimates of bomb craters showed considerably more pronounced increments in uncertainty. As a matte of fact, increments in uncertainty obtain only for *S*s who changed their estimates. These results were analyzed by comparing the trial blocks immediately preceding and immediately succeeding stimulus and feedback change. For the purposes of this analysis the results of all change conditions were combined, and the mean SU scores for the two critical trial blocks are shown in Table 6. The analysis of variance for these results is shown in Table 7. It is clear that while the high changers increased their uncertainty, low changers continued to decrease it. It is also clear that the likelihood of change is higher for *S*s with high original uncertainty, although the analysis failed to generate a significant interaction.

TABLE 6

MEAN SU SCORES BEFORE AND AFTER STIMULUS AND FEEDBACK CHANGE FOR HIGH AND LOW CHANGERS

	Mean SU Scores	
	Before Change	After Change
High Changers	101.9	108.5
Low Changers	81.3	71.5

TABLE 7

ANALYSIS OF VARIANCE FOR DATA IN TABLE 6

Source	*df*	*MS*	*F*
High vs. Low Changers (C)	1	29,756	2.16
*S*s in same group	70	13,798	
Trial blocks (B)	1	97	<1.00
B × C	1	2,418	8.17**
Pooled *S*s by B	70	296	

** $p < .01$.

This result supports in part the implication of the unfreezing hypothesis, which holds that the unfreezing process occurs only when necessary, that is, when the stability prior to change is relatively high. The differences in uncertainty increments between the b group and the c group also give some support to this implication, although neither of these findings is very convincing. A more direct test is obtained by comparing *S*s whose uncertainty level before change induction was relatively high with those whose uncertainty level was low. Such a comparison was made and the results are shown in Figure 6. By means of the SU scores on the trial block immediately preceding change induction, *S*s in all change con-

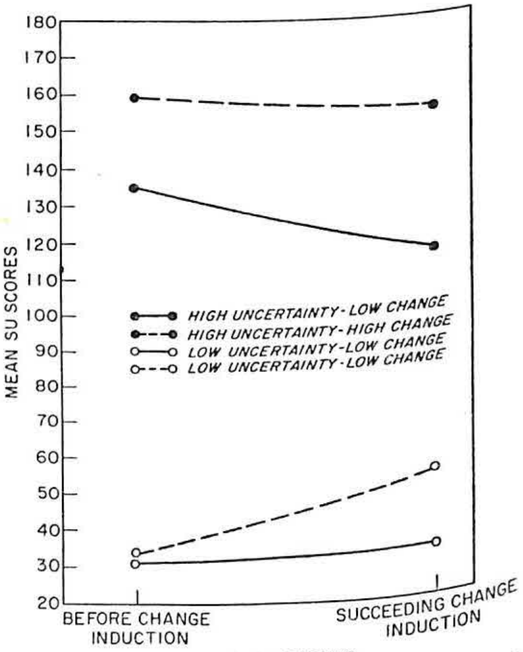

FIG. 6. Effects of cognitive change on uncertainty (SU scores) for subjects with initially high and low uncertainty.

TABLE 8
ANALYSIS OF VARIANCE FOR DATA IN FIGURE 6

Source	*df*	*MS*	*F*
Uncertainty (U)	1	387,195	45.14***
High vs. Low Changers (C)	1	17,009	1.98
U × C	1	3,239	<1.00
*S*s in same group	68	8,577	
Trial blocks (T)	1	13	<1.00
T × U	1	5,053	12.86***
T × C	1	2,154	5.48*
T × U × C	1	142	<1.00
Pooled *S*s × T	68	393	

* $p < .05$.
*** $p < .001$.

ditions were split at the median, and each half was further divided at the median in terms of change in the direction of slide change or feedback change. Thus four groups of *S*s were obtained; High Uncertainty–High Change, High Uncertainty–Low Change; Low Uncertainty–High Change, and Low Uncertainty–Low Change. The analysis was performed on the uncertainty changes from the trial block immediately preceding to the trial block immediately succeeding change induction, and the analysis of variance is shown in Table 8. It is evident from Figure 6 that increment in uncertainty occurs only for *S*s whose initial uncertainty was low. *S*s with high initial uncertainty show decrements whether they changed their estimates or not. The interaction between Trial Blocks and Uncertainty is highly significant. It is also clear that in the Low Uncertainty group there is some increment in uncertainty for *S*s who changed their estimates relatively less. The previous result relating increments in uncertainty to changes in estimates is again confirmed. In the Low Uncertainty group, those *S*s who changed their estimates show a considerably more pronounced uncertainty increment than those who did not. As predicted, in the High Uncertainty group there is no increment in uncertainty for high changers, but *S*s above the median amount of change reduced their uncertainty to a much lesser degree than *S*s below median amount of change. The interaction between Trial Blocks and Change is significant at the .05 level. The strong main effect of uncertainty is of course due to the arbitrary division of the population at the median.

SUMMARY

The purpose of this study was to examine the role of uncertainty in cognitive change. It was hypothesized that when a person must for some reason change his judgments or opinions his uncertainty should increase. Moreover, it was predicted that uncertainty increments are necessary if cognitive change is to occur, but only when the initial uncertainty is relatively low. Ninety-six subjects were given the task of estimating the number of bomb craters in an aerial photograph in 18 trials involving brief exposures of the stimulus. Judgmental change was induced by either changing the stimulus or changing feedback. Objective and subjective measures of uncertainty were obtained. The results lead to the following conclusions:

1. Repeated exposures of the stimulus result in progressive uncertainty decrements.
2. When requirement for change is introduced the level of uncertainty rises.
3. The extent of the uncertainty increments depends on the uncertainty level before change. With high uncertainty no increments in uncertainty occur; with low uncertainty there are pronounced increments.
4. Uncertainty increments occur only for persons who change their judgments.

REFERENCES

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

LEVITT, E. E. The Water-Jar Einstellung Test as a measure of rigidity. *Psychol. Bull.*, 1956, **53**, 347–370.

LEWIN, K. Group decision and social change. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley, (Eds.), *Readings in Social Psychology*. (Rev. ed.) New York: Holt, 1952. Pp. 459–473.

ZAJONC, R. B., & BURNSTEIN, E. Cognitive behavior under uncertainty. *Amer. Psychologist*, 1957, **12**, 388. (Abstract)

(Received August 17, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 176-180

LEVELING-SHARPENING AND MEMORY ORGANIZATION[1]

PHILIP S. HOLZMAN AND RILEY W. GARDNER
Menninger Foundation, Topeka, Kansas

The terms, "leveling" and "sharpening," have been used to describe systematic changes in memory organization (Carmichael, Hogan, & Walter, 1932) and the fate of rumors (Allport & Postman, 1947). Omission of inconsistencies, condensation of elements, and general simplification of recalled stories have all been called "leveling." The tendency of a detail of the original structure of a story to dominate its memory organization has been called "sharpening." Wulf (1922) reported analogous processes in the reproduction of geometric figures.

Our use of the terms, "leveling" and "sharpening," shares certain features of these meanings, but implies more. We employ the terms to describe opposite poles of a dimensional principle of cognitive control concerning the degree of assimilation between perceptual processes and memory traces (Köhler, 1923; Lauenstein, 1932). The concept of cognitive control is based on the observation that individual differences in cognitive behaviors often reflect habitual, but not necessarily conscious, strategies of adaptation, and the further observation (see Klein, 1954) that the enduring response dispositions represented by such strategies serve to *mediate* the effects of drives upon cognitive behavior. Cognitive controls such as those we have referred to as the leveling-sharpening dimension are presumed to be structural principles of personality organization that give a particular cast to an individual's cognitive behavior.

Generalized leveling-sharpening tendencies have been explored and defined in several prior studies. The assimilation hypothesis just mentioned led to successful prediction of significant relationships between performance in a laboratory test of assimilation-proneness (the Schematizing Test) and the degree of assimilation shown in visual, auditory, and kinesthetic time error (Holzman, 1954; Holzman & Klein, 1954). Extreme leveling has also been shown to be associated with reliance upon repression as a preferred mode of defense (Gardner, Holzman, Klein, Linton, & Spence, 1959; Holzman & Gardner, 1959) and degree of distortion experienced through aniseikonic lenses (Gardner, et al., 1959).

The most complete study of the leveling-sharpening control principle to date was done by Gardner et al. (1959). In that study, leveling-sharpening emerged as a clear-cut factor for women, and included test scores used as measures of leveling-sharpening in previous studies, among other measures. This factor was described as pertaining "to differentiation in memory organization as a function of the extent to which successive stimuli assimilate to each other." Apparently because of relatively great assimilation among new percepts and relevant memory traces formed earlier, subjects (*S*s) with high factor scores (levelers) formed undifferentiated memory organizations in several tests. *S*s with low factor scores (sharpeners) formed relatively differentiated memory organizations, presumably by virtue of minimal assimilation. A definitive test of the differences in memory schemata in levelers and sharpeners was not, however, undertaken. The present study represents a first such test.

In the earlier studies cited, the criterion of leveling-sharpening has been the Schematizing Test. A complete description of the test may be found in Gardner et al. (1959). Briefly, *S*s are asked to judge in inches the sizes of 150 squares successively projected on a screen. The 14 stimuli range in size from approximately 1 to 14 inches on a side and are projected in a prescribed order that is unknown to the *S*s. The five smallest squares are individually presented (3″ exposure, 8″ interval) in ascending order, then in two random orders. Without *S*'s knowledge, the smallest square is then dropped and one just larger than any seen previously is added. This group of five squares is also presented in ascending order and in two random orders. Subtraction of the smallest square and addition of a larger square is continued in this fashion until all 14 stimuli have been projected.

---

[1] This study was made possible by Research Grant M-1182 from the National Institutes of Health, Public Health Service. We are indebted to our colleague, Richard Siegal, for his help with this study, and to Gardner Murphy, Robert R. Holt, and David Rapaport for their helpful criticisms.

Several facets of individual difference are apparent in the Schematizing Test. In the present study, we shall deal with only one of these: the accuracy with which *S*s rank the 150 squares presented (computed in terms of ranking accuracy within each series of 15 judgments, a procedure fully described by Gardner et al.). Accurate intraserial ranking requires continuous comparison of the newly exposed square with those presented earlier. If new percepts and memories of earlier squares undergo relatively great change by virtue of assimilation, intraserial ranking accuracy is notably impaired. Thus, discrimination of the relative sizes of temporarily juxtaposed stimuli may be blurred by assimilation.

If our interpretation of the leveling-sharpening principle is correct, *S*s at different points along this continuum should differ predictably in their recall of much earlier experiences, as well as in their registration of temporal sequences of stimulation in the laboratory. Requiring *S*s to retell a story learned long ago, but not purposefully committed to memory, appears to provide a test of the hypothesis that such memory schemata differ in levelers and sharpeners. We would expect young adults who are levelers to have relatively few correct memories of such a story as "The Pied Piper" and to show evidence of relatively great loss and modification of the over-all structure of the story. This general hypothesis is, of course, difficult to put to adequate test: some *S*s may have heard the story more frequently during childhood than others, some may have reheard the story more recently than others, etc. These features of their prior experience can obviously not be controlled. Control over other extraneous determinants of recall may have been provided, however, by the following facts: (*a*) the age range of our 41 female *S*s was only 18 to 22, (*b*) their socioeconomic backgrounds were remarkably similar, (*c*) they were all sorority members at one small midwestern university. If age, for example, had been allowed to vary more, the relationships reported here might easily have been obscured.

METHOD

The Schematizing Test was administered to 41 *S*s under the standard conditions described by Gardner et al. For purposes of the present study *S*s were divided into groups of 16 levelers and 25 sharpeners on the basis of the mean percentage accuracy of intraserial ranking for five previous samples totaling 175 female *S*s.

In a different testing session, each *S* was asked to tell the story of "The Pied Piper."[2] (Those who reported never having heard it were eliminated.) The stories were recorded.

For purposes of scoring number of elements correct, the story was arbitrarily divided into 11 thematic units:

1. Mention of the town of Hamlin.
2. The town was infested with rats.
3. The townspeople were unable to rid themselves of the rats.
4. The Pied Piper, however, offered to exterminate the rats.
5. If a specified sum of money were paid to him.
6. When he played his pipe, the rats followed him.
7. To a body of water where they drowned.
8. The mayor (or townspeople) refused to pay the Piper.
9. The Piper then played his pipe, which, this time, attracted all the children of the town.
10. The Piper led them out of town into a mountain.
11. Only the lame boy remained behind, because he was unable to keep up with the rest.

Since the story presumably represented "old" memory for most of our *S*s, it seemed advisable to scrutinize the preservation of the main outlines of the story, as well as the correct and incorrect responses at the more microscopic level. The two major sections of the story are linked by a key theme: the Pied Piper's leading the children away as *retribution* for the town fathers' refusal to pay him. Each protocol was thus scored for presence or absence of this key theme, without which ellipsis, condensation, or omission of parts of the over-all organization must occur.

In addition, the following error scores were obtained by the two experimenters (*E*s): number of new elements; number of recombinations; number of fragmentations; number of confabulations, i.e., the number of thematic units associatively elaborated. Because of the simplicity of the scoring scheme, there were only minor differences in scoring. These were easily reconciled to provide a single set of scores.

RESULTS

Quantitative

As anticipated, sharpeners were superior to levelers in their recall of "The Pied Piper." The mean numbers of elements correct were

---

[2] The use of this story was suggested by David Rapaport.

5.6 for sharpeners, 3.3 for levelers, $t = 2.24$, $p < .02$, one-tailed test.[3]

The two groups also differed in retention of the key "retribution" theme. Only 26% of the levelers retained this theme, whereas 58% of the sharpeners' stories were anchored in this central organizational feature of the story.

The stories told by sharpeners were more than half again as long as those told by levelers. Mean number of typed lines in the recall protocols were: sharpeners, 12.9, levelers, 8.3; $t = 1.86$, $p < .05$, one-tailed test. The two groups were not significantly different in number of new elements, number of recombinations of elements, number of fragmentations, or number of confabulations. The fact that sharpeners produced much longer stories without more errors means, however, that many of their additional elements—those not included in our scoring scheme, which dealt only with correct elements and gross errors—were "sensible" additions that bore a reasonable relationship to the original organization of the story. Since the number-correct score was based only on elements common to various versions of "The Pied Piper," many of these additional elements may represent the particular versions originally heard or read by individual sharpeners.

Qualitative

Qualitative differences in the performances of levelers and sharpeners were perhaps even more impressive than the quantifiable differences just described.

Consider first a performance by one of the levelers that is unique in respect to the core "retribution" theme. This *S* demonstrated her loss of the over-all organization of the story by presenting *two* stories—representing the last and then the first parts of the original story. The loss of the key retribution theme showed itself in a fragmentation of the over-all structure of the story such that *S* apparently remembered it as two separate and unreconcilable entities.

> Well there was once a man in this village who played a very wonderful, I guess it was a magic flute, and he could walk through the town and all the children would hear him and be so entranced that they would follow him along to listen to him play. And so he, I guess, went through the villages collecting all the children. I [pause], let's see. . . . Oh . . . I think *this* is the, I think *this* is the Pied Piper. This village was full of mice and rats and so in order to get rid of them they loved music and so there was this man that could play a flute and so he went through the village and played the flute and out of the houses came the mice and rats and they followed him and he led them into the ocean and in that way the town got rid of all the rodents [laughs]. [*E* asks *S* about first story.] Oh, I can't answer the first one because I can't remember what he did with the children, the first story, the one I started and didn't finish.

In the following story, told by a sharpener, the sequence and transition are much more clearly preserved.

> The people of Hamlin were over-run by rats and they had tried all sorts of remedies to get rid of the rats and nothing worked. So, one day, they were discussing the rat problem as usual, and this man came to town and the mayor had just previously offered a bag of gold to anyone who could rid the town of rats, so this man came to town. His name was Pied Piper and he told the mayor that he would rid the town of rats for the bag of gold. So, he played on his little pipe and all the rats followed him out—came down from the houses and followed him down the street and took them out of town and they disappeared. He came back for his gold, and the mayor refused to give it to him on the grounds that he never promised it. So Pied Piper told him he would be sorry if he didn't give him the gold, that he would take away all the children. Well, the mayor just laughed at him because he didn't think he could do it. So, he still refused to give him the gold, so the Pied Piper played the tune again and all the children followed him out of the town and he took them out into the country where some mountains—supposedly, the mountains opened up and all the children all went inside to this paradise, and the children—the mountains closed and the children were never seen again.

Vagueness was also notable in levelers' recall. Often only an indistinct general impression of the story could be recovered, with important details missing. The following story by a leveler seems to highlight these qualities.

> Well, I probably didn't know any facts about it but, well, let's see, didn't he get rid of the mice or rats or something from the town? I don't remember. But as I remember, they, they were, they were trying to get rid of them or something, and he by playing his pipe, or whatever it was, musical instrument, anyway, by doing that, he drove them out of town, but that's all I've ever known. [*E*: Do you remember anything else

---

[3] In a large group of which the present *S*s were a part, the Pearson correlations between ACE scores and leveling-sharpening were not significantly different from zero. Intelligence as traditionally measured, then, is not a significant factor in the differences between levelers and sharpeners on recall of "The Pied Piper."

at all?] *S*: Unh uh. [*E*: Do you have the feeling that there's more to the story?] *S*: Oh, yes, but I probably just don't know.

DISCUSSION

Theories of learning and remembering may be arranged along a continuum of greater to lesser influence of "personality" factors on these cognitive functions. In the special theories of Ebbinghaus, for example, the role of general personality factors enters not at all. Associationistic theories of learning in which the law of effect plays a crucial role emphasize the part played by affective factors in learning and retention. Gestalt theories of learning and remembering elucidate chiefly the ways in which structural properties of the material to be learned determine registration and recall. Others, principally Bartlett (1932), have refined our awareness that memories themselves have organizational properties shaped in part by affects. Psychoanalysis has attempted to take into account the central role of affects in memory in the theory of repression. A complete review and thoughtful critique of these issues has been provided by Rapaport (1950). It may fairly be said that, in general, conceptions of the role of personality in memory have been limited to special effects, such as needs, affects, and specific sets. The concentration on such influences upon memory seems to emerge from an implicit assumption that personality factors are external influences on the learning or recall processes, factors that act either to enhance or to interfere with memory.

Theories of learning and remembering have usually been constructed without reference to individual differences in personality *organization* that shape both registration and recall. The general laws propounded thus far bypass the possibility that Person A and Person B may learn and remember in strikingly different ways, although when their performances are taken together with those of others, the group averages may show conformity to such laws.[4] Recall is a complex process in which details lose or maintain their place within the context of the structural organization of the story as this structure was shaped at the time(s) of the person's exposure to the story.

The cognitive control principles referred to earlier are concepts employed to describe and account for some of the major dimensions along which *personal constants* in thinking, perceiving, remembering, etc., fall. These controls are assumed to represent aspects of the individual's personality organization that determine the nature and quality of such cognitive functions. Leveling-sharpening—a dimension of consistent individual differences in susceptibility to assimilation (process-trace interaction)—is a principle of cognitive organization that affects the registration, and hence recall, of stimuli in a wide variety of situations, here shown to include a test of recall of "The Pied Piper" story. Leveling-sharpening thus seems to be apparent in the organization of old memories, as well as in the organization of memories registered in a current laboratory experiment. This finding implies that leveling-sharpening is a generalized and enduring aspect of cognitive organization. *S*s at the leveling end of this dimension of cognitive organization formed memory schemata for a story heard much earlier that have the following major characteristics: loss of the over-all structure of the story, loss of elements as a product of "fusion" in the process of registration, relatively large numbers of elements (considering the number of recallable elements available) contaminated or changed by interaction with other memory-elements-to-be during the registration process.

Repression, one of the special "personality" factors affecting memory, has received considerable clinical attention from psychoanalysts. In previous studies (Gardner, et al., 1959; Holzman & Gardner, 1959), we showed that persons who employ repression as a principal defense mechanism tend to be levelers, i.e., that a major characteristic of the memory apparatus is discernible in both defensive and adaptive uses of memory. In the light of the earlier findings, results of the present study also suggest that a generalized tendency to form undifferentiated memory schemata provides a necessary but not a sufficient condition for the unavailability to intentional verbal expression of ideas that, in psychoanalytic terms, have become drive-cathected. Such generalized clogging in the recall of ideas by levelers is consistent with McGeoch's (1929) demonstration that the higher the degree of the original learning, the less the interference in the reproduction of the material. A further investigation is indicated to determine whether, among levelers, repression occurs in a diffuse fashion, while instances of repression among sharpeners are more discrete and confined to special circumstances.

---

[4] I. H. Paul's (1959) study of individual differences in remembering, however, represents a novel and fruitful approach to understanding individual styles of learning and recall, particularly as these styles affect the "importation" of various kinds of extraneous material in recall.

Experiments now in progress are designed to determine whether or not levelers' undifferentiated initial memory organizations are subject to relatively great "spontaneous" changes after original registration is completed. Although the present study points to a link between registration in the Schematizing Test, earlier registration of "The Pied Piper" story, and subsequent registration of events *S* experienced as related to the story (at some level of conceptual organization), our results may also be due to post-registration changes that result primarily from lack of differentiation in levelers' memories. A further possibility, pointed to by Gardner et al. (1959), must also be put to the test: that proneness to assimilative phenomena is a function of the intensity with which persons attend to new experiences. Persons who characteristically "cathect" new experiences weakly may be susceptible to assimilation (i.e., may be levelers) because of vagueness and instability in their original percepts.

REFERENCES

ALLPORT, G. W., & POSTMAN, L. *The psychology of rumor.* New York: Holt, 1947.

BARTLETT, F. C. *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge Univer. Press, 1932.

CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.

GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, HARRIET B., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Issues*, 1959, **1**, No. 4.

HOLZMAN, P. S. The relation of assimilation tendencies in visual, auditory, and kinesthetic time-error to cognitive attitudes of leveling and sharpening. *J. Pers.*, 1954, **22**, 375–394.

HOLZMAN, P. S., & GARDNER, R. W. Leveling and repression. *J. abnorm. soc. Psychol.*, 1959, **59**, 151–155.

HOLZMAN, P. S., & KLEIN, G. S. Cognitive system-principles of leveling and sharpening: Individual differences in assimilation effects in visual time-error. *J. Psychol.*, 1954, **37**, 105–122.

KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1954. Pp. 224–274.

KÖHLER, W. Zur Theorie des Sukzessivvergleichs und der Zeitfehler. *Psychol. Forsch.*, 1923, **4**, 115–175.

LAUENSTEIN, O. Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler. *Psychol. Forsch.*, 1932, **17**, 130–177.

MCGEOCH, J. A. The influence of degree of learning upon retroactive inhibition. *Amer. J. Psychol.*, 1929, **41**, 252–262.

PAUL, I. H. Studies in remembering: The reproduction of connected and extended verbal material. *Psychol. Issues*, 1959, **1**, No. 2.

RAPAPORT, D. *Emotions and memory.* New York: International Univer. Press, 1950.

WULF, F. Über die Veränderung von Vorstellungen (Gedächtnis und Gestalt). *Psychol. Forsch.*, 1922, **1**, 333–373.

(Early publication received April 4, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 181-189

THE EFFECT OF THREAT UPON INTERPERSONAL BARGAINING

MORTON DEUTSCH AND ROBERT M. KRAUSS
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey

A BARGAIN is defined in *Webster's Unabridged Dictionary* as "an agreement between parties settling what each shall give and receive in a transaction between them"; it is further specified that a bargain is "an agreement or compact viewed as advantageous or the reverse." When the term "agreement" is broadened to include tacit, informal agreements as well as explicit agreements, it is evident that bargains and the processes involved in arriving at bargains ("bargaining") are pervasive characteristics of social life.

The definition of bargain fits under sociological definitions of the term "social norm." In this light, the experimental study of the bargaining process and of bargaining outcomes provides a means for the laboratory study of the development of certain types of social norms. But unlike many other types of social situations, bargaining situations have certain distinctive features that make it relevant to consider the conditions that determine whether or not a social norm will develop as well as those that determine the nature of the social norm if it develops. Bargaining situations highlight the possibility that, even where cooperation would be mutally advantageous, shared purposes may not develop, agreement may not be reached, and interaction may be regulated antagonistically rather than normatively.

The essential features of a bargaining situation exist when:

1. Both parties perceive that there is the possibility of reaching an agreement in which each party would be better off, or no worse off, than if no agreement were reached.

2. Both parties perceive that there is more than one such agreement that could be reached.

3. Both parties perceive each other to have conflicting preferences or opposed interests with regard to the different agreements that might be reached.

Everyday examples of bargaining include such situations as: the buyer-seller relationship when the price is not fixed, the husband and wife who want to spend an evening out together but have conflicting preferences about where to go, union-management negotiations, drivers who meet at an intersection when there is no clear right of way, disarmament negotiations.

In terms of our prior conceptualization of cooperation and competition (Deutsch, 1949) bargaining is thus a situation in which the participants have mixed motives toward one another: on the one hand, each has interest in cooperating so that they reach an agreement; on the other hand, they have competitive interests concerning the nature of the agreement they reach. In effect, to reach agreement the cooperative interest of the bargainers must be strong enough to overcome their competitive interests. However, agreement is not only contingent upon the *motivational* balances of cooperative to competitive interests but also upon the situational and *cognitive* factors which facilitate or hinder the recognition or invention of a bargaining agreement that reduces the opposition of interest and enhances the mutuality of interest.[1]

These considerations lead to the formulation of two general, closely related propositions about the likelihood that a bargaining agreement will be reached.

1. Bargainers are more likely to reach an agreement, the stronger are their cooperative interests in comparison with their competitive interests.

2. Bargainers are more likely to reach an agreement, the more resources they have available for recognizing or inventing potential bargaining agreements and for communicating to one another once a potential agreement has been recognized or invented.

From these two basic propositions and additional hypotheses concerning conditions that determine the strengths of the cooperative and competitive interests and the amount of available resources, we believe it is possible to explain the ease or difficulty of arriving at a bargaining agreement. We shall not present a full statement of these hypotheses here but turn instead to a description of an experiment that relates to Proposition 1.

---

[1] Schelling in a series of stimulating papers on bargaining (1957, 1958) has also stressed the "mixed motive" character of bargaining situations and has analyzed some of the cognitive factors which determine agreements.

The experiment was concerned with the effect of the availability of threat upon bargaining in a two-person experimental bargaining game.[2] Threat is defined as the expression of an intention to do something detrimental to the interests of another. Our experiment was guided by two assumptions about threat:

1. If there is a conflict of interest and one person is able to threaten the other, he will tend to use the threat in an attempt to force the other person to yield. This tendency should be stronger, the more irreconcilable the conflict is perceived to be.

2. If a person uses threat in an attempt to intimidate another, the threatened person (if he considers himself to be of equal or superior status) would feel hostility toward the threatener and tend to respond with counterthreat and/or increased resistance to yielding. We qualify this assumption by stating that the tendency to resist should be greater, the greater the perceived probability and magnitude of detriment to the other and the less the perceived probability and magnitude of detriment to the potential resister from the anticipated resistance to yielding.

The second assumption is based upon the view that when resistance is not seen to be suicidal or useless, to allow oneself to be intimidated, particularly by someone who does not have the right to expect deferential behavior, is to suffer a loss of social face and, hence, of self-esteem; and that the culturally defined way of maintaining self-esteem in the face of attempted intimidation is to engage in a contest for supremacy vis-à-vis the power to intimidate or, minimally, to resist intimidation. Thus, in effect, the use of threat (and if it is available to be used, there will be a tendency to use it) should strengthen the competitive interests of the bargainers in relationship to one another by introducing or enhancing the competitive struggle for self-esteem. Hence, from Proposition 1, it follows that the availability of a means of threat should make it more difficult for the bargainers to reach agreement (providing that the threatened person has some means of resisting the threat). The preceding statement is relevant to the comparison of both of our experimental conditions of threat, bilateral and unilateral (described below), with our experimental condition of nonthreat. We hypothesize that a bargaining agreement is more likely to be achieved when neither party can threaten the other, than when one or both parties can threaten the other.

Consider now the situations of bilateral threat and unilateral threat. For several reasons, a situation of bilateral threat is probably less conducive to agreement than is a condition of unilateral threat. First, the sheer likelihood that a threat will be made is greater when two people rather than one have the means of making the threat. Secondly, once a threat is made in the bilateral case it is likely to evoke counterthreat. Withdrawal of threat in the face of counterthreat probably involves more loss of face (for reasons analogous to those discussed in relation to yielding to intimidation) than does withdrawal of threat in the face of resistance to threat. Finally, in the unilateral case, although the person without the threat potential can resist and not yield to the threat, his position vis-à-vis the other is not so strong as the position of the threatened person in the bilateral case. In the unilateral case, the threatened person may have a worse outcome than the other whether he resists or yields; while in the bilateral case, the threatened person is sure to have a worse outcome if he yields but he may insure that he does not have a worse outcome if he does not yield.

METHOD

Procedure

Subjects (*S*s) were asked to imagine that they were in charge of a trucking company, carrying merchandise over a road to a destination. For each trip completed they made \$.60, minus their operating expenses. Operating expenses were calculated at the rate of one cent per second. So, for example, if it took 37 seconds to complete a particular trip, the player's profit would be \$.60 − \$.37 or a net profit of \$.23 for that particular trip.

Each *S* was assigned a name, Acme or Bolt. As the "road map" (see Figure 1) indicates, both players start from separate points and go to separate destinations. At one point their paths cross. This is the section of road labeled "one lane road," which is only one lane wide, so that two trucks, heading in opposite directions, could not pass each other. If one backs up

---

[2] The game was conceived and originated by M. Deutsch; R. M. Krauss designed and constructed the apparatus employed in the experiment.

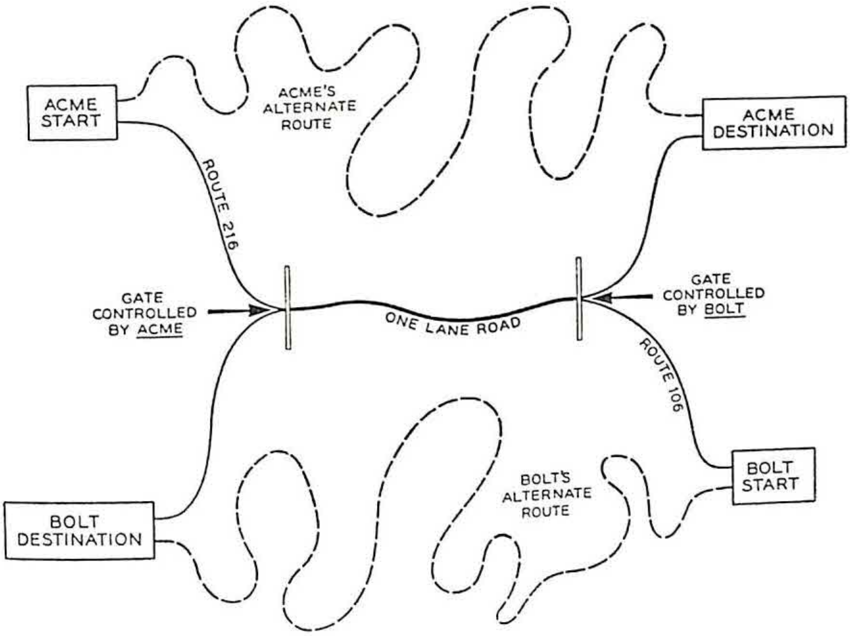

FIG. 1. Subject's road map.

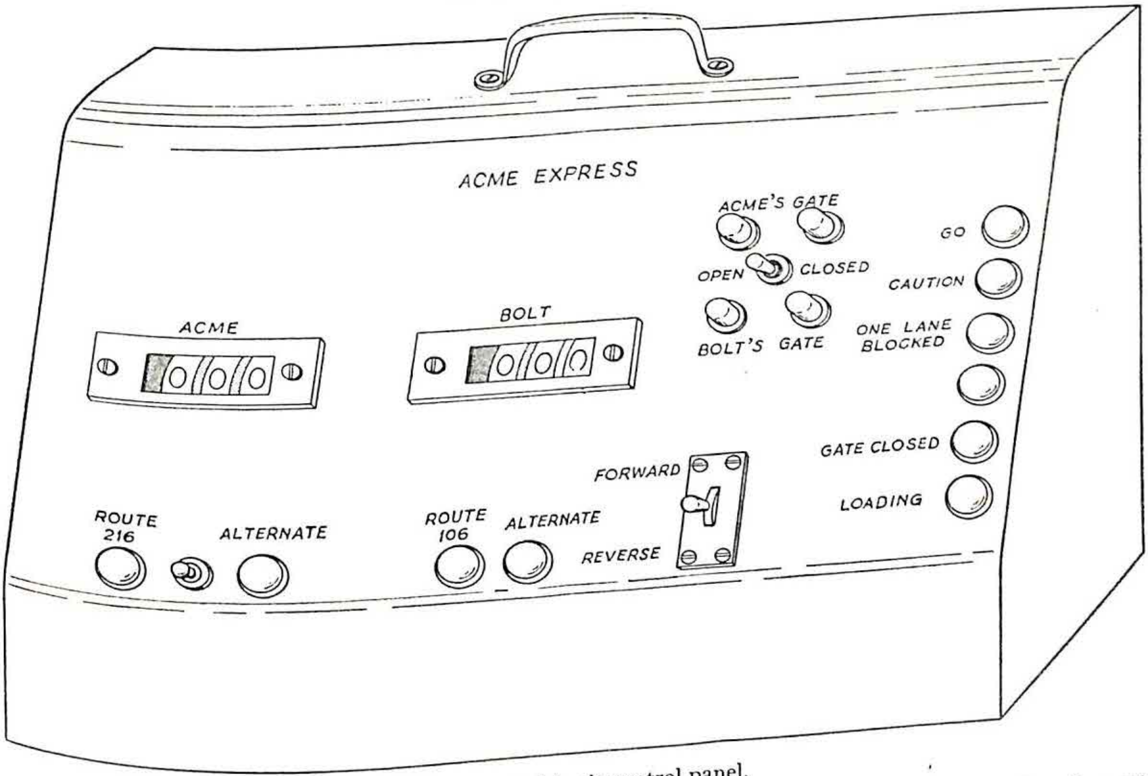

FIG. 2. Subject's control panel.

the other can go forward, or both can back up, or both can sit there head-on without moving.

There is another way for each *S* to reach the destination on the map, labeled the "alternate route." The two players' paths do not cross on this route, but the alternate is 56% longer than the main route. *S*s were told that they could expect to lose at least $.10 each time they used the alternate route.

At either end of the one-lane section there is a gate that is under the control of the player to whose starting point it is closest. By closing the gate, one player can prevent the other from traveling over that section of the main route. The use of the gate provides the threat potential in this game. In the bilateral threat potential condition (Two Gates) both players had gates under their control. In a second condition of unilateral threat (One Gate) Acme had control of a gate but Bolt did not. In a third condition (No Gates) neither player controlled a gate.

Ss played the game seated in separate booths placed so that they could not see each other but could see the experimenter (*E*). Each *S* had a "control panel" mounted on a 12″ × 18″ × 12″ sloping-front cabinet (see Figure 2). The apparatus consisted essentially of a reversible impulse counter that was pulsed by a recycling timer. When the *S* wanted to move her truck forward she threw a key that closed a circuit pulsing the "add" coil of the impulse counter mounted on her control panel. As the counter cumulated, *S* was able to determine her "position" by relating the number on her counter to reference numbers that had been written in on her road map. Similarly, when she wished to reverse, she would throw a switch that activated the "subtract" coil of her counter, thus subtracting from the total on the counter each time the timer cycled.

S's counter was connected in parallel to counters on the other *S*'s panel and on *E*'s panel. Thus each player had two counters on her panel, one representing her own position and the other representing the other player's. Provision was made in construction of the apparatus to permit cutting the other player's counter out of the circuit, so that each *S* knew only the position of her own truck. This was done in the present experiment. Experiments now in progress are studying the effects of knowledge of the other person's position and other aspects of interpersonal communication upon the bargaining process.

The only time one player definitely knew the other player's position was when they had met head-on on the one-way section of road. This was indicated by a traffic light mounted on the panel. When this light was on, neither player could move forward unless the other moved back. The gates were controlled by toggle switches and panel-mounted indicator lights showed, for both *Ss*, whether each gate was open or closed.

The following "rules of the game" were stated to the *Ss*:

1. A player who started out on one route and wished to switch to the other route could only do so after first reversing and going back to the start position. Direct transfer from one route to the other was not permitted except at the start position.
2. In the conditions where *Ss* had gates, they were permitted to close the gates no matter where they were on the main route, so long as they were on the main route (i.e., they were not permitted to close the gate while on the alternate route or after having reached their destinations). However, *Ss* were permitted to open their gates at any point in the game.

Ss were taken through a number of practice exercises to familiarize them with the game. In the first trial they were made to meet head-on on the one-lane path; Acme was then told to back up until she was just off the one-lane path and Bolt was told to go forward. After Bolt had gone through the one-lane path, Acme was told to go forward. Each continued going forward until each arrived at her destination. The second practice trial was the same as the first except that Bolt rather than Acme backed up after meeting head-on. In the next practice trial, one of the players was made to wait just before the one-way path while the other traversed it and then was allowed to continue. In the next practice trial, one player was made to take the alternate route and the other was made to take the main route. Finally, in the bilateral and unilateral threat conditions the use of the gate was illustrated (by having the player get on the main route, close the gate, and then go back and take the alternate route). The *Ss* were told explicitly, with emphasis, that they did *not* have to use the gate. Before each trial in the game the gate or gates were in the open position.

The instructions stressed an individualistic motivational orientation. *Ss* were told to try to earn as much money for themselves as possible and to have no interest in whether the other player made money or lost money. They were given $4.00 in poker chips to represent their working capital and told that after each trial they would be given "money" if they made a profit or that "money" would be taken from them if they lost (i.e., took more than 60 seconds to complete their trip). The profit or loss of each *S* was announced so that both *Ss* could hear the announcement after each trial. Each pair of *Ss* played a total of 20 trials; on all trials, they started off together. In other words each trial presented a repetition of the same bargaining problem. In cases where *Ss* lost their working capital before the 20 trials were completed, additional chips were given them. *Ss* were aware that their monetary winnings and losses were to be imaginary and that no money would change hands as a result of the experiment.

Subjects

Sixteen pairs of *Ss* were used in each of the three experimental conditions. The *Ss* were female clerical and supervisory personnel of the New Jersey Bell Telephone Company who volunteered to participate during their working day.[3] Their ages ranged from 20 to 39, with a mean of 26.2. All were naive to the purpose of the experiment. By staggering the arrival times and choosing girls from different locations, we were able to insure that the *Ss* did not know with whom they were playing.

Data Recorded

Several types of data were collected. We obtained a record of the profit or loss of each *S* on each trial. We also obtained a detailed recording of the actions taken by each *S* during the course of a trial. For this purpose, we used an Esterline-Angus model AW Operations Recorder which enabled us to obtain a "log" of each move each *S* made during the game (e.g., whether and when she took the main or alternate route; when she went forward, backward, or remained still; when she closed and opened the gate; when she arrived at her destination).

RESULTS[4]

The best single measure of the difficulty experienced by the bargainers in reaching an agreement is the sum of each pair's profits (or losses) on a given trial. The higher the sum of the payoffs to the two players on a given trial, the less time it took them to arrive at a procedure for sharing the one-lane path of the main route. (It was, of course, possible for one or both of the players to decide to take the alternate route so as to avoid a protracted stalemate during the process of bargaining. This, however, always resulted in at least a $.20 smaller joint payoff if only one player took the alternate route, than an optimally arrived at agreement concerning the use of the one-way path.) Figure 3 presents the medians of the summed payoffs (i.e., Acme's plus Bolt's) for all pairs in each of the three experimental conditions over the 20 trials.[5] These striking results indicate that agreement was least difficult to arrive at in the no threat condition, was more difficult to arrive at in the unilateral threat condition, and exceedingly difficult or impossible to arrive at in the bilateral threat condition (see also Table 1).

---

[3] We are indebted to the New Jersey Bell Telephone Company for their cooperation in providing *Ss* and facilities for the experiment.

[4] We are indebted to M. J. R. Healy for suggestions concerning the statistical analysis of our data.

Examination of Figure 3 suggests that learning occurred during the 20 trials: the summed payoffs for pairs of *S*s tend to improve as the number of trials increases. This suggestion is confirmed by an analysis of variance of the slopes for the summed payoffs[6] over the 20 trials for each of the 16 pairs in each of the 3 experimental treatments. The results of this analysis indicate that the slopes are significantly greater than zero for the unilateral threat ($p < .01$) and the no threat ($p < .02$) conditions; for the bilateral threat condition, the slope does not reach statistical significance ($.10 < p < .20$). The data indicate that the pairs in the no threat condition started off at a fairly high level but, even so, showed some improvement over the 20 trials; the pairs in the unilateral threat condition started off low and, having considerable opportunity for improvement, used their opportunity; the pairs in the bilateral threat condition, on the other hand, did not benefit markedly from repeated trials.

Figure 4 compares Acme's median profit in the three experimental conditions over the 20 trials; while Figure 5 compares Bolt's profit in

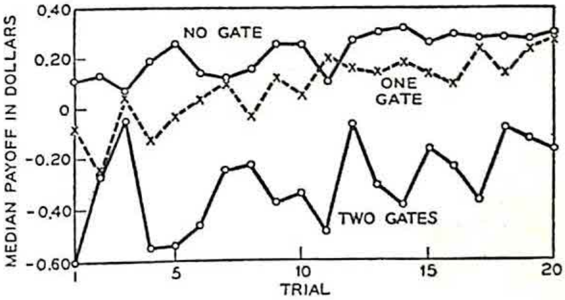

FIG. 3. Median joint payoff (Acme + Bolt) over trials.

the three conditions. (In the unilateral threat condition, it was Acme who controlled a gate and Bolt who did not.) Bolt's as well as Acme's outcome is somewhat better in the no threat condition than in the unilateral threat condition; Acme's, as well as Bolt's, outcome is clearly worst in the bilateral threat condition (see Table 1 also). However, Figure 6 reveals that Acme does somewhat better than Bolt in the unilateral condition. Thus, if threat-potential exists within a bargaining relationship it is better to possess it oneself than to have the other party possess it. However, it is even better for neither party to possess it. Moreover, Figure 5 shows that Bolt is better off not having than having a gate even when Acme has a gate: Bolt tends to do better in the unilateral threat condition than in the bilateral threat condition.

The size of the absolute discrepancy between the payoffs of the two players in each pair provides a measure of the confusion or difficulty in predicting what the other player was going to do. Thus, a large absolute discrepancy might indicate that after one player had gone through the one-way path and left it open, the other player continued to wait; or it might indicate that one player continued to wait at a closed gate hoping the other player would open it quickly but the other player did not; etc. Figure 7 indicates that the discrepancy between players in the no threat condition is initially small and remains small for the 20 trials. For the players in both the bilateral and unilateral threat conditions, the discrepancy is initially relatively larger; but it decreases more noticeably in the unilateral threat condition by the tenth trial and, therefore, is consistently smaller than in the bilateral condition.

By way of concrete illustration, we present a

---

[5] Medians are used in graphic presentation of our results because the wide variability of means makes inspection cumbersome.

[6] A logarithmic transformation of the summed payoffs on each trial for each pair was made before computing the slopes for a given pair.

TABLE 1

MEAN PAYOFFS SUMMATED OVER THE TWENTY TRIALS

Variable	Means			Statistical Comparisons: p values[a]			
	(1) No Threat	(2) Unilateral Threat	(3) Bilateral Threat	Overall	(1)vs. (2)	(1)vs.(3)	(2)vs.(3)
Summed Payoffs (Acme + Bolt)	203.31	−405.88	−875.12	.01	.01	.01	.05
Acme's Payoff	122.44	−118.56	−406.56	.01	.10	.01	.05
Bolt's Payoff	80.88	−287.31	−468.56	.01	.01	.01	.20
Absolute Differences in Payoff (A − B)	125.94	294.75	315.25	.05	.05	.01	*ns*

[a] Evaluation of the significance of overall variation between conditions is based on an F test with 2 and 45 *df*. Comparisons between treatments are based on a two-tailed t test.

synopsis of the game for one pair in each of three experimental treatments.

No Threat Condition

Trial 1. The players met in the center of the one-way section. After some back-and-forth movement Bolt reversed to the end of the one-way section, allowing Acme to pass through, and then proceeded forward herself.

Trial 2. They again met at the center of the one-way path. This time, after moving back and forth deadlocked for some time, Bolt reversed to "start" and took the alternate route to her destination, thus leaving Acme free to go through on the main route.

Trial 3. The players again met at the center of the one-way path. This time, however, Acme reversed to the beginning of the path, allowing Bolt to go through to her destination. Then Acme was able to proceed forward on the main route.

Trial 5. Both players elected to take the alternate route to their destinations.

Trial 7. Both players took the main route and met in the center. They waited, deadlocked, for a considerable time. Then Acme reversed to the end of the one-way path allowing Bolt to go through, then proceeded through to her destination.

Trials 10–20. Acme and Bolt fall into a pattern of alternating who is to go first on the one-way section. There is no deviation from this pattern.

The only other pattern that emerges in this condition is one in which one player dominates the other. That is, one player consistently goes first on the one-way section and the other player consistently yields.

Unilateral Threat Condition

Trial 1. Both players took the main route and met in the center of it. Acme immediately closed the gate, reversed to "start," and took the alternate route to her destination. Bolt waited for a few seconds, at the closed gate, then reversed and took the alternate route.

Trial 2. Both players took the main route and met in the center. After moving back and forth deadlocked for about 15 seconds, Bolt reversed to the beginning of the one-way path, allowed Acme to pass, and then proceeded forward to her destination.

Trial 3. Both players started out on the main route, meeting in the center. After moving back and forth deadlocked for a while, Acme closed her gate, reversed to "start," and took the alternate route. Bolt, meanwhile, waited at the closed gate. When Acme arrived at her destination she opened the gate, and Bolt went through to complete her trip.

Trial 5. Both players took the main route, meeting at the center of the one-way section. Acme immediately closed her gate, reversed, and took the alternate route. Bolt waited at the gate for about 10 seconds, then reversed and took the alternate route to her destination.

Trial 10. Both players took the main route and met in the center. Acme closed her gate, reversed, and took the alternate route. Bolt remained waiting at the closed gate. After Acme arrived at her destination, she opened the gate and Bolt completed her trip.

Trial 15. Acme took the main route to her destination and Bolt took the alternate route.

Trials 17–20. Both players took the main route and met in the center. Bolt waited a few seconds, then reversed to the end of the one-way section allowing Acme to go through. Then Bolt proceeded forward to her destination.

Other typical patterns that developed in this experimental condition included an alternating pattern similar to that described in the no threat condition, a dominating pattern in which Bolt would select the alternate route leaving Acme free to use the main route unobstructed, and a pattern in which Acme would close her gate and then take the alternate route, also forcing Bolt to take the alternate route.

Bilateral Threat Condition

Trial 1. Acme took the main route and Bolt took the alternate route.

Trial 2. Both players took the main route and met head-on. Bolt closed her gate. Acme waited a few seconds, then closed her gate, reversed to "start," then went forward again to the closed gate. Acme reversed and took the alternate route. Bolt again reversed, then started on the alternate route. Acme opened her gate and Bolt reversed to "start" and went to her destination on the main route.

Trial 3. Acme took the alternate route to her destination. Bolt took the main route and closed her gate before entering the one-way section.

Trial 5. Both players took the main route and met head-on. After about 10 seconds spent backing up and going forward, Acme closed her gate, reversed, and took the alternate route. After waiting a few seconds, Bolt did the same.

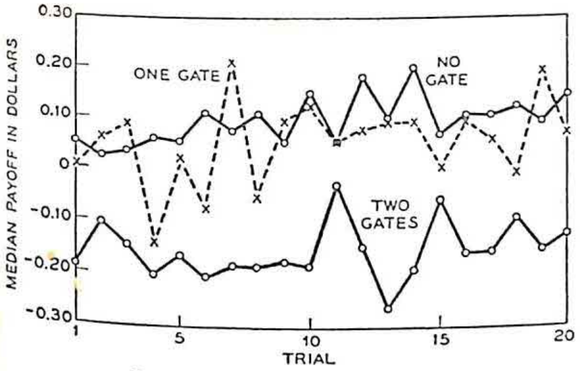

FIG. 4. Acme's median payoff.

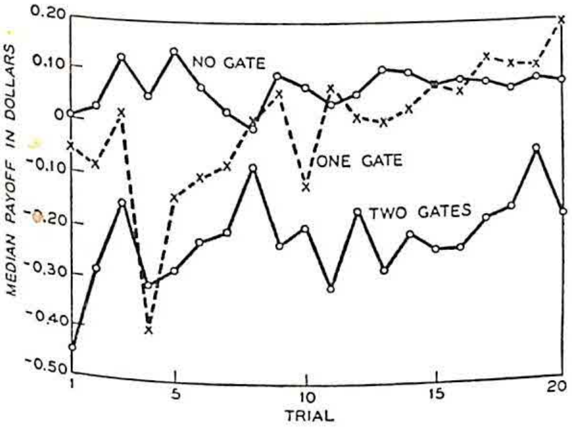

FIG. 5. Bolt's median payoff.

Trials 8–10. Both players started out on the main route, immediately closed their gates, reversed to "start," and took the alternate route to their destinations.

Trial 15. Both players started out on the main route and met head-on. After some jockeying for position, Acme closed her gate, reversed, and took the alternate route to her destination. After waiting at the gate for a few seconds, Bolt reversed to "start" and took the alternate route to her destination.

Trials 19–20. Both players started out on the main route, immediately closed their gates, reversed to "start," and took the alternate routes to their destinations.

Other patterns that emerged in the bilateral threat condition included alternating first use of the one-way section, one player's dominating the other on first use of the one-way section, and another dominating pattern in which one player consistently took the main route while the other consistently took the alternate route.

DISCUSSION

From our view of bargaining as a situation in which both cooperative and competitive tendencies are present and acting upon the individual, it is relevant to inquire as to the conditions under which a stable agreement of any form develops. However, implicit in most economic models of bargaining (e.g., Stone, 1958; Zeuthen, 1930) is the assumption that

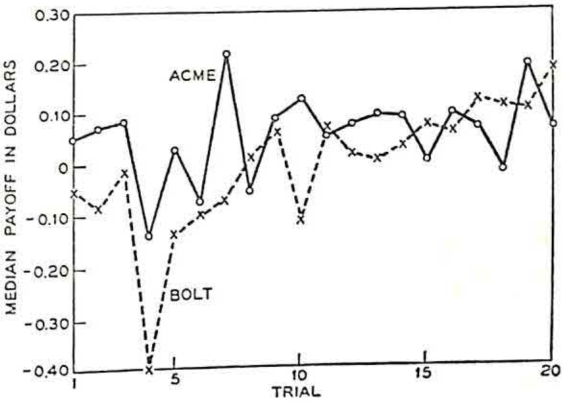

FIG. 6. Acme's and Bolt's median payoffs in unilateral threat condition.

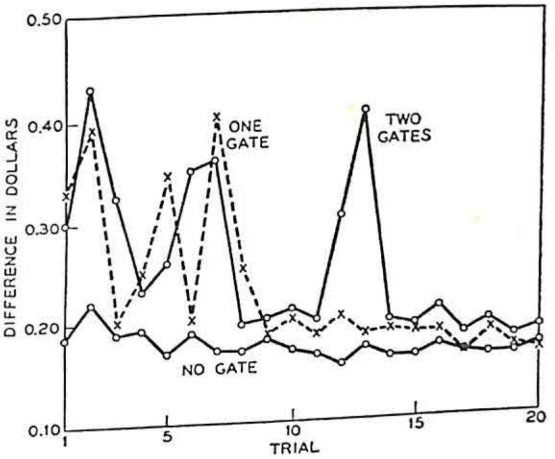

FIG. 7. Median absolute differences in payoff.

the cooperative interests of the bargainers are sufficiently strong to insure that some form of mutually satisfactory agreement will be reached. For this reason, such models have focused upon the form of the agreement reached by the bargainers. Siegel and Fouraker (1960) report a series of bargaining experiments quite different in structure from ours in which only one of many pairs of *S*s were unable to reach agreement. Siegel and Fouraker explain this rather startling result as follows:

> Apparently the disruptive forces which lead to the rupture of some negotiations were at least partially controlled in our sessions . . . .
>
> Some negotiations collapse when one party becomes incensed at the other, and henceforth strives to maximize his opponent's displeasure rather than his own satisfaction . . . . Since it is difficult to transmit insults by means of quantitative bids, such disequilibrating behavior was not induced in the present studies. If subjects were allowed more latitude in their communications and interactions, the possibility of an affront-offense-punitive behavior sequence might be increased (p. 100). (Quoted by permission of McGraw-Hill)

In our experimental bargaining situation, the availability of threat clearly made it more difficult for bargainers to reach a mutually profitable agreement. These results, we believe, reflect psychological tendencies that are not confined to our bargaining situation: the tendency to use threat (if the means for threatening is available) in an attempt to force the other person to yield, when the other is seen as obstructing one's path; the tendency to respond with counterthreat or increased resistance to attempts at intimidation. How general are these tendencies? What conditions are likely to elicit them? Answers to these questions are necessary before our results can be generalized to other situations.

Dollard, Doob, Miller, Mowrer, and Sears (1939) have cited a variety of evidence to support the view that aggression (i.e., the use of threat) is a common reaction to a person who is seen as the agent of frustration. There seems to be little reason to doubt that the use of threat is a frequent reaction to interpersonal impasses. However, everyday observation indicates that threat does not inevitably occur when there is an interpersonal impasse. We would speculate that it is most likely to occur: when the threatener has no positive interest in the other person's welfare (he is either egocentrically or competitively related to the other); when the threatener believes that the other has no positive interest in his welfare; and when the threatener anticipates either that his threat will be effective or, if ineffective, will not worsen his situation because he expects the worst to happen if he does not use his threat. We suggest that these conditions were operative in our experiment; *S*s were either egocentrically or competitively oriented to one another[7] and they felt that they would not be worse off by the use of threat.

Everyday observation suggests that the tendency to respond with counterthreat or increased resistance to attempts at intimidation is also a common occurrence. We believe that introducing threat into a bargaining situation affects the meaning of yielding. Although we have no data to support this interpretation directly, we will attempt to justify it on the basis of some additional assumptions.

Goffman (1955) has pointed out the pervasive significance of "face" in the maintenance of the social order. In this view, self-esteem is a socially validated system that grows out of the acceptance by others of the claim for deference, prestige, and recognition that a person presents in his behavior toward others. Since the rejection of such a claim would be perceived (by the recipient) as directed against his self-esteem, he must react against it rather than accept it in order to maintain the integrity of his self-esteem system.

One may view the behavior of our *S*s as an attempt to make claims upon the other, an attempt to develop a set of shared expectations as to what each was entitled to. Why then did the *S*s' reactions differ so markedly as a function of the availability of threat? The explanation lies, we believe, in the cultural interpretation of yielding (to a peer or subordinate) under duress, as compared to giving in without duress. The former, we believe, is perceived as a negatively valued form of behavior, with negative implications for the self-image of the person who so behaves. At least partly, this is so because the locus of causality is perceived to be outside the person's voluntary control. No such evaluation, however, need be placed on the behavior of one who "gives in" in a situation where no threat or duress is a factor. Rather, we should expect the culturally defined evaluation of such a person's behavior to be one of "reasonableness" or "maturity," because the source of the individual's behavior is perceived to lie within his own control.

Our discussion so far has suggested that the psychological factors which operate in our experimental bargaining situation are to be found in many real-life bargaining situations. However, it is well to recognize some unique features of our experimental game. First, the bargainers had no opportunity to communicate verbally with one another. Prior research on the role of communication in trust (Deutsch 1958, 1960; Loomis, 1959) suggests that the opportunity for communication would have made reaching an agreement easier for individualistically-oriented bargainers. This same research (Deutsch, 1960) indicates, however, that communication may not be effective between competitively oriented bargainers. This possibility was expressed spontaneously by a number of our *S*s in a post-game interview.

---

[7] A post-experimental questionnaire indicated that, in all three experimental conditions, the *S*s were most strongly motivated to win money, next most strongly motivated to do better than the other player, next most motivated to "have fun," and were very little or not at all motivated to help the other player.

Another characteristic of our bargaining game is that the passage of time, without coming to an agreement, is costly to the players. There are, of course, bargaining situations in which lack of agreement may simply preserve the *status quo* without any worsening of the bargainers' respective situations. This is the case in the typical bilateral monopoly case, where the buyer and seller are unable to agree upon a price (e.g., see Siegel & Fouraker, 1960). In other sorts of bargaining situations, however, (e.g., labor-management negotiations during a strike, internation negotiations during an expensive cold war) the passage of time may play an important role. In our experiment, we received the impression that the meaning of time changed as time passed without the bargainers reaching an agreement. Initially, the passage of time seemed to place the players under pressure to come to an agreement before their costs mounted sufficiently to destroy their profit. With the continued passage of time, however, their mounting losses strengthened their resolution not to yield to the other player. They comment: "I've lost so much, I'll be damned if I give in now. At least I'll have the satisfaction of doing better than she does." The mounting losses and continued deadlock seemed to change the game from a mixed motive into a predominantly competitive situation.

It is, of course, hazardous to generalize from a laboratory experiment to the complex problems of the real world. But our experiment and the theoretical ideas underlying it can perhaps serve to emphasize some notions which, otherwise, have an intrinsic plausibility. In brief, these are that there is more safety in cooperative than in competitive coexistence, that it is dangerous for bargainers to have weapons, and that it is possibly even more dangerous for a bargainer to have the capacity to retaliate in kind than not to have this capacity when the other bargainer has a weapon. This last statement assumes that the one who yields has more of his values preserved by accepting the agreement preferred by the other than by extended conflict. Of course, in some bargaining situations in the real world, the loss incurred by yielding may exceed the losses due to extended conflict.

SUMMARY

The nature of bargaining situations was discussed. Two general propositions about the conditions affecting the likelihood of a bargaining agreement were presented. The effects of the availability of threat upon interpersonal bargaining were investigated experimentally in a two-person bargaining game. Three experimental conditions were employed: no threat (neither player could threaten the other), unilateral threat (only one of the players had a means of threat available to her), and bilateral threat (both players could threaten each other). The results indicated that the difficulty in reaching an agreement and the amount of (imaginary) money lost, individually as well as collectively, was greatest in the bilateral and next greatest in the unilateral threat condition. Only in the no threat condition did the players make an overall profit. In the unilateral threat condition, the player with the threat capability did better than the player without the threat capability. However, comparing the bilateral and unilateral threat conditions, the results also indicate that when facing a player who had threat capability one was better off *not* having than having the capacity to retaliate in kind.

REFERENCES

DEUTSCH, M. A theory of cooperation and competition. *Hum. Relat.*, 1949, **2**, 129–152.

DEUTSCH, M. Trust and suspicion. *J. conflict Resolut.*, 1958, **2**, 265–279.

DEUTSCH, M. The effect of motivational orientation upon trust and suspicion. *Hum. Relat.*, 1960, **13**, 123–140.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. H. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

GOFFMAN, E. On face-work. *Psychiatry*, 1955, **18**, 213–231.

LOOMIS, J. L. Communication, the development of trust and cooperative behavior. *Hum. Relat.*, 1959, **12**, 305–315.

SCHELLING, T. C. Bargaining, communication and limited war. *J. conflict Resolut.*, 1957, **1**, 19–38.

SCHELLING, T. C. The strategy of conflict: Prospectus for the reorientation of game theory. *J. conflict Resolut.*, 1958, **2**, 203–264.

SIEGEL, S., & FOURAKER, L. E. *Bargaining and group decision making.* New York: McGraw-Hill, 1960.

STONE, J. J. An experiment in bargaining games. *Econometrica*, 1958, **26**, 286–296.

ZEUTHEN, F. *Problems of monopoly and economic warfare.* London: Routledge, 1930.

(Early publication received July 20, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 190–206

VERBAL CONDITIONING OF TWO RESPONSE CLASSES:

SOME METHODOLOGICAL CONSIDERATIONS[1]

JOSEPH D. MATARAZZO, GEORGE SASLOW, AND E. NELSON PAREIS[2]

University of Oregon Medical School

VERBAL conditioning has recently aroused considerable interest (Krasner, 1958; Salzinger, 1959). The research here being reported led to the investigation of several parameters of verbal conditioning which seemed to need additional study: (*a*) response class conditioned (Are *human* responses and *plural* responses equally conditionable?); (*b*) influence of different experimenters (Can one *E* cross-validate the findings of another *E*?); (*c*) influence on verbal conditioning of such subject variables as age, sex, education, intelligence (WAIS vocabulary level), Taylor anxiety level, and total time to complete the experiment; (*d*) the role of *S*'s awareness of the purpose of the experiment on his conditioning or nonconditioning performance; (*e*) the reliability of independent observer-scorers in the verbal conditioning studies here being reported; and (*f*) the influence of *S*'s base rate on his conditioning score.

METHOD

The studies were conducted on a floor housing the Division of Medical Psychology in the University of Oregon Medical School Hospital. Two soundproofed rooms, separated by a one-way mirror, and equipped with a microphone in one room and an amplifier and earphones in the observers' room were used for the experiments.

Subjects. All *S*s were paid undergraduate volunteers from Portland State College.[3] Their median age was 20, the mean was 23.5, and the range from 18 to 47. Mean educational level was sophomore year (Grade 13.4), with a range from freshman through senior year. There were 80 *S*s in all in the first four studies (20 in each), yielding a sex distribution of 48 females and 32 males. This ratio of 1.5 females to each male was roughly maintained in each of the four studies. All these *S* characteristics were obtained in the two additional studies (*N* of 20 and *N* of 10) which were conducted following the first four main studies.

Experimenters. The two male *E*s in the first four studies were a 33-year-old psychologist and a 52-year-old psychiatrist. Each *E* saw a different *S* in an odd-even design, thus yielding an original ($N = 20$) and replication ($N = 20$) experiment for the Plurals ($N = 40$) study, and an original ($N = 20$) and replication ($N = 20$) experiment for the Humans ($N = 40$) study. To further check the results of the two Plurals studies, the same two *E*s studied 20 additional *S*s (10 each, odd-even design) in a fifth study, and a third *E* (a 36-year-old male psychologist) studied 10 additional *S*s in a sixth study. All three *E*s wore their usual white hospital coats[4] during the experiment. In order to check the reliability of each *E*'s own recording and scoring of *S*'s responses during each experiment, the remaining two *E*s served as observers (*O*s) on the other side of the one-way mirror. Each *O* recorded and scored *S*'s responses independently. Thus it was possible to check *E*'s reliability of response-class judgment against each of the two *O*s and to check one *O* against the other.

Procedure. *S*s were selected from psychology classes in which it was known that neither the lectures nor the textbooks dealt with verbal conditioning. *S*s were asked to sign up "for a study on language." Each *S* was met by the designated *E* and, after being seated in the experimental room, was given these instructions:

> We are doing a study on language and how people use words. Your job today will be very simple. I want you to say words—any words at all will do, and there is no time limit—but they must be words; no sentences and no numbers. There are two people behind the one-way mirror making more elaborate counts of your words than the tally I will make. Your job, then, is simply to say words. If you have any questions about what you are to do, please ask me now because once we start I will not be able to answer your questions [pause for questions with repetition of instructions given where necessary]. All right, if you have no (additional) questions, go ahead. Say all the words you can until I tell you to stop. Go!

The conditioning procedure was as follows: In Period I (Operant or Baseline period), for *S*'s first 100 words, *E* said nothing. In Period II (Conditioning), for *S*'s next 200 words, *E* said "Good" or "That's good" every time *S* gave a word in the response class being conditioned. In Period III (Extinction), for the next 100 words, *E* again said nothing.

Because a pilot experiment had demonstrated that the first introduction of the reinforcement "Good" or "That's good," coming as it did after 100 or more words where *E* had said nothing, occasioned in some *S*s either startle or a response such as "Am I finished?" it was decided to introduce one such reinforcement by E early in the experiment. Thus, unknown to S, all the studies reported here used a pre-experiment period of 10 words, following the *third word* of which E said "Good" for every S, no matter what the word. Thus, S was introduced early to a reinforcing response by E. Because this introductory reinforcement was given only once, and *randomly* distributed across verbal response classes of all types, it was assumed that it would have little, if any, systematic effect on the main study. Experience demonstrated that this procedure reduced considerably, or removed completely, the previously described unwanted effects otherwise seen the first time E introduced the reinforcement in Period II. No rules were given E about the use of "Good" versus "That's good" as reinforcement. The two variations were introduced in order to make E's verbal behavior sound less stereotyped. E was asked to use either form spontaneously and in whatever order his own spontaneity-set suggested.

---

[1] Paper read at American Psychological Association, Cincinnati, September 1959. This research was supported by a research grant (M-1938) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] Now at Portland State College.

[3] We are grateful to Lester F. Beck and his staff for their generous assistance in the recruitment of these *S*s.

[4] The influence of this "status" symbol may be important in view of studies (Kanfer, 1958, pp. 443–444; Krasner, 1958, p. 152; Salzinger, 1959, p. 80; Verplanck, 1955, p. 598) suggesting a relationship between *S*'s conditionability and personal characteristics of the examiner (in this case, *E*'s implied level of prestige or social status).

Two response classes were reinforced in separate experiments: Plural Nouns and Human Responses. The former class was defined as any plural noun ending in the letter "s." This response class is relatively easy to score, although by agreement it was decided to score such words as, for example, "pants" as a *nonplural* when the context suggested a single pair. Human Responses were considered any word which clearly and unambiguously designated a person. Thus, "mother," "brother," "architect," "neighbor," etc. were included, but not "American" or "Australian," since the latter could be adjectives. Exceptions to this latter were scored when the context clearly suggested a person: e.g., "Englishman," "Frenchman," "American," "Scandinavian," "Spaniard." Such words as "necktie," "hands," etc. were not reinforced; only clear references to persons were.

Recording by both Es and each of the two Os was facilitated considerably by the use of a prepared recording sheet. This consisted of a single sheet of plain white paper on which single upright (tally) lines, in groups of five, were typed, ten groups of five or fifty to a line. The first 100, the next 200, and the last 100 were separated as Periods I, II, and III, respectively. E and the two Os drew a mark through each tally line for every word not in the response category being conditioned, and drew a circle through each tally line for every word in the appropriate category. Speed and accuracy of scoring thus were markedly enhanced.

Immediately following the conditioning procedure, each S was introduced to a different examiner, who individually administered the following tests in this order: (*a*) a specially prepared, open-ended, *written* questionnaire for assessing S's "level of awareness" of the purpose of the experiment; (*b*) the vocabulary subtest of the Wechsler Adult Intelligence Scale (Wechsler, 1955), and the Taylor Manifest Anxiety Scale (Taylor, 1953).

RESULTS

As shown graphically in Figure 1 and in tabular form in the top half of Table 1, for the two Plurals studies, neither E_1 nor E_2 produced any significant conditioning in Period II by the use of a reinforcement following each statement by S of a Plural Noun in that period. With the two groups of Ss each giving their respective Es approximately 16% Plural Nouns in the free or operant period, neither group increased their output of Plurals significantly with the introduction of the verbal reinforcement in Period II. As a matter of fact, E_2's group showed a slight (nonsignificant) decrease in Period II. Since no conditioning was evident in Period II, this precluded examination for extinction in Period III.

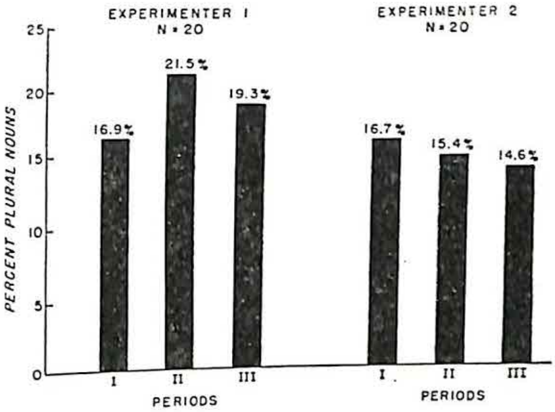

FIG. 1. Percent frequency of plural nouns in three periods for original experiment (E_1) and replication (E_2).

That neither E was able to condition Plural Nouns in these two studies is remarkable in view of the fact that seven out of the seven published studies of Plural Nouns reviewed by Krasner (1958, p. 160) yielded positive results. That the reinforcement used may explain our negative results is suggested by the fact that no one of these seven positive studies utilized "Good" as a reinforcement. Used instead were "mmm-hmm" (3 studies), a light (2 studies), a buzzer (1 study), and "huh-uh" (1 study). While six of these seven studies showed a significant increase in Plural Nouns during reinforcement, Greenspoon (1955) reported that the use of "huh-uh" by E produced a statistically significant *decrease* in Ss' production of Plural Nouns. (Greenspoon's Ss did show an *increase* in Plural Nouns to "mmm-hmm" as a reinforcer.) Table 1 and Figure 1 show that the trend (nonsignificant) of the data for our E_2 is in the same direction (decrease) as that of Greenspoon's, even though, following a lead from the study by Mandler and Kaplan (1956), our "Good" might be considered

TABLE 1

PERCENTAGE OF WORDS GIVEN IN RESPONSE CLASS BEING REINFORCED IN EACH OF THREE PERIODS

		Periods			
		I	II	III	*F* test
Study 1—PLURAL NOUNS					
Original Study (E_1) ($N = 20$)	*M*	16.9	21.5	19.3	.86
	SD	10.97	20.61	19.88	
	R	2–45	2–92	0–77	
Replication (E_2) ($N = 20$)	*M*	16.7	15.4	14.6	.50
	SD	14.08	10.94	10.64	
	R	1–58	0–40	1–38	
Study 2—HUMANS					
Original Study (E_1) ($N = 20$)	*M*	6.7	30.9	20.7	10.60***
	SD	4.45	23.58	28.33	
	R	1–15	6–87	3–100	
Replication (E_2) ($N = 20$)	*M*	5.0	20.2	10.7	9.41***
	SD	4.14	20.83	8.98	
	R	0–16	3–79	0–37	

*** $p < .001$.

a positive reinforcer and his "huh-uh" a negative reinforcer. This decrease for E_2 may be merely a chance occurrence. Further research will be necessary to determine the reasons for our inability to condition our *S*s on Plural Nouns in the presence of so many successful studies in the literature. It is worth mentioning, however, that colleagues from other parts of the country have reported in informal conversation similar lacks of success in conditioning Plural Nouns. Such experiences, plus our own negative results (Figure 1), would suggest that either the type of verbal reinforcement used or some other presumably identifiable variable is responsible for the different results. Examination and discussion of some of the variables which might have led to our negative results with Plural Nouns are presented later in this paper.

In contrast to the results with Plural Nouns, Table 1 and Figure 2 show that the same two *E*s were able successfully to condition the *S*s to the response class of Humans. Both their groups of 20 *S*s started off with a similar percentage (6.7 and 5.0) of Human Responses in the free-operant Period I. The mean percentage values across the three periods for both *E*s (6.7, 30.9, 20.7; and 5.0, 20.2, 10.7 for E_1 and E_2, respectively) show clear evidence of conditioning in Period II and extinction in Period III. These differences across the three periods are highly significant ($p < .001$) for each *E* by either the *F* test for correlated means (McNemar, 1955, pp. 290–296) as shown in Table 1, or (not shown) the nonparametric Friedman's two-way analysis of variance by ranks (Siegel, 1956, pp. 166–173). The latter test was employed as a check on the *F* test when a variant of Bartlett's *F* test (Edwards, 1954, pp. 327–328) on the standard deviations shown in Table 1 was significant for the Humans (but not for the Plurals) studies of both *E*s, suggesting nonhomogeneity of variance in the three periods, for this response class. (Similar nonparametric analysis of the data for the Plurals study yielded a nonsignificant *F*, again indicating no conditioning for this response class.)

Thus, as E_2 cross-validated the negative results of E_1 with Plural Nouns, he also cross-validated the positive results with Human Responses. This fact of independent cross-validation of the results by a second *E* using *S*s from the same population (see Krasner, 1958, p. 152) would suggest that the results obtained in this study are highly reliable. The additional fact (Table 1) that the base rates (obtained by the two *E*s from independent groups of 20 *S*s each) are remarkably similar for both response classes is a second indication of reliability. Such alternate-groups and alternate-*E*s

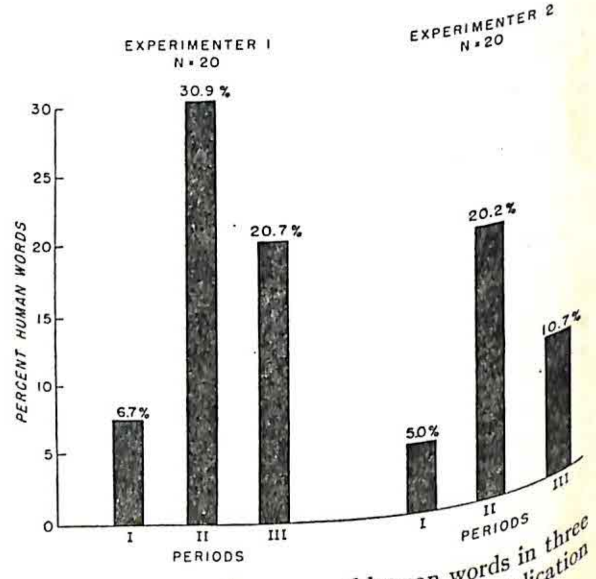

FIG. 2. Percent frequency of human words in three periods for original experiment (E_1) and replication (E_2).

TABLE 2

PERIOD COMPARISONS OF PERCENTAGE OF WORDS IN EACH OF TWO RESPONSE CLASSES

	Periods			
	I	II	III	*F*
EXPERIMENTER 1				
Plural Nouns	16.9	21.5	19.3	.86
Human	6.7	30.9	20.7	10.60***
F	14.74***	1.75	.03	
EXPERIMENTER 2				
Plural Nouns	16.7	15.4	14.6	.50
Humans	5.0	20.2	10.7	9.41***
F	12.66***	.82	.77	

*** $p < .001$.

reliability is a useful substitute for the test-retest reliability check in studies where necessary procedures (like the inquiry relating to "awareness") may make retest with the same *S*s difficult if not impossible.

Table 2 provides a statistical answer to the question: Are different response classes emitted equally frequently by *S*s in the baseline or free-operant period? It is clear from the results in this table that, while the baselines for Plural Nouns (16.9 and 16.7) and for Human Responses (6.7 and 5.0) were remarkably similar across *E*s, the percentage of Human Responses produced by the *S*s differed significantly from the percentage of Plural Nouns produced by the groups in the Plurals Study. As shown in Table 2 (reading down), analysis by conventional *F* test of the differences (16.9 versus 6.7% and 16.7 versus 5%, for E_1 and E_2, respectively) in the Period I base level production of these two response classes by the *S*s seen by each *E* (i.e., analysis across response class) was highly significant ($p < .001$). Interestingly, despite the evidence (reading across in Table 2) for conditioning in one response class and not the other, unlike the case with Period I, the differences in Periods II and III were not significant across the two response classes (reading down, see nonsignificant *F* values for Periods II and III for both *E*s in Table 2). That is, by use of reinforcement it was possible, in Period II, to bring the frequency of emission of one response class (Humans) up to the frequency of emission of the other (Plurals), even though the base rates were demonstrably different initially.

Despite the fact that the four studies shown in Table 2 used different *S*s instead of the same *S*s for assessing Period I base rates for these two response classes, the uniformity of the findings across *E*s (16.9 versus 6.7 and 16.7 versus 5.03, respectively) is further evidence of a reliable difference in response-class base rates. In addition, inspection of the means for these two response categories published by other investigators suggests that the means here are of the same order of magnitude as in other college populations (Mandler & Kaplan, 1956).

Thus, it can be concluded that Plural Nouns seem to have a lower threshold or higher emission rate than do Human Responses in the free-operant phase of verbal conditioning experiments of the type described here. If investigations in this general area continue their present proliferation, it may very well become possible to determine the free-operant frequencies of various response categories of different populations of *S*s. Wilson and Verplanck (1956) already have shown that adverbs, as a response class, have a very low operant level relative to Plural Nouns.

While significant, the results shown in Tables 1 and 2 and in Figures 1 and 2 merely represent averages or group trends. The phenomenon of "verbal conditioning" would appear to have so many theoretical as well as practical implications that it seemed wise to do a rather extensive analysis of the performance of individuals within these groups. We asked such questions as these: (*a*) Is it possible that a few *S*s in the Plural Nouns group, by starting off with large Period I base rate scores and then earning much lower Period II scores, may have influenced the results unduly by masking a majority phenomenon? (*b*) Granted that the group trends suggest that *S*'s response of Humans, as a response class, can be conditioned and extinguished by *E*'s behavior, how general is this phenomenon? Do all *S*s show it, or are the positive findings the result of a few *S*s who show extreme conditioning?

In order to examine the performance of individuals, it was necessary to define a criterion of conditioning or nonconditioning for each *S*. In the absence of any other criteria, we used

TABLE 3

OBSERVED AND EXPECTED FREQUENCIES OF INDIVIDUAL SUBJECTS FALLING IN EACH OF FOUR POSSIBLE PATTERNS ACROSS PERIODS[a]

	N	A		B		C		D		Chi square
PLURAL NOUNS										
E_1	20	5	(5)[b]	5	(5)	4	(5)	6	(5)	.20
E_2	20	4	(5)	6	(5)	4	(5)	6	(5)	.20
Both *E*s	40	9	(10)	11	(10)	8	(10)	12	(10)	.50
HUMANS										
E_1	20	1	(5)	18	(5)	1	(5)	0	(5)	40.20**
E_2	20	5	(5)	13	(5)	0	(5)	2	(5)	16.60**
Both *E*s	40	6	(10)	31	(10)	1	(10)	2	(10)	56.10**

[a] The following four patterns for the three periods were considered the most probable:

A. Base–Up–Up
B. Base–Up–Down (apparent conditioners)
C. Base–Down–Up
D. Base–Down–Down

[b] Numbers enclosed in parentheses are expected values.

** $p < .01$.

as a first step the crude index that any *S* who, relative to his own baseline, increased his production in Period II and decreased it in Period III was a "conditioner," i.e., there was a contingency between his behavior and that of the *E*. In order to account for the fact that, independent of any effect produced by *E*'s reinforcement, any given *S* could show such a pattern by chance alone, we found it necessary to define the frequency of "chance" or *apparent conditionability*. This we established as one *S* out of 4, or 25%, by the following reasoning (see bottom of Table 3). For any individual *S*, given a baseline or operant value of any magnitude (from 0 to 100%), his performance under reinforcement in Period II can be either up or down. Thus, taking Period I as a base, two directions are possible to a given individual by chance alone. With Period II as reference point, *S* also has two possibilities open to him in Period III, up or down. With such a wide possible range (from 0 to 100 in Periods I, II, and III) and our own empirical data to inspect, we decided a tie in Period II relative to Period I, or a tie in Period III relative to Period II, was so infrequent as to make these two cases not of the same order of chance magnitude as the other four cases just described. Thus, we considered the following four cases as the most probable patterns, independent of what *E* is doing: A. Base–Up–Up; B. Base–Up–Down (chance, or apparent conditioner); C. Base–Down–Up; and D. Base–Down–Down. An exception to this model would be the *S* who obtained a score of 100% in either Period I or Period II. This individual could then only go down (or tie) in the subsequent period. Since empirically no *S* showed such a pattern, we discarded consideration of this possible but unlikely exception.

If by chance alone, then, any given *S* showed our Class B, or apparent conditioning pattern, one would expect five such *S*s in each of our four studies employing 20 *S*s. Table 3 gives the obtained and expected numbers of *S* who showed each of the four patterns in the Plurals and the Humans study. The results are shown individually for each *E* and for the total group for each response class.

Inspection of the obtained and expected frequencies and of the chi square values (corrected for continuity) shows that the results yield merely chance relationships for the Plurals. For each of the two *E*s, by chance alone, 5 out of their 20 *S*s would be expected to show the B pattern. The obtained frequency of 5 and 6 for E_1 and E_2, respectively, confirms the nonsignificant group results shown in Tables 1 and 2 for this response clase. The obtained frequencies of approximately 5 for the other three patterns (A, C, and D) for each of the two *E*s also indicate chance frequencies, thus empirically confirming our earlier stated hypothesis that, in the absence of an effect due to *E*'s behavior, each of the four patterns should produce a chance frequency of 5 out of 20.

As also shown in Table 3, the analysis by individuals for the Humans study yields 18 out of 20 *S*s showing the B pattern for E_1 ($p < .01$) and 13 out of 20 ($p < .01$) for E_2 where, again, only 5 each would be expected by chance alone. The difference between the obtained frequencies (13 and 18) for the two *E*s was not significant (chi square of .80). Thus, 31 out of 40 *S*s in the two Humans studies showed conditioning. With appropriate allowance for chance (*N* of 10 out of 40), this still leaves 21 out of 40, or one *S* out of every two, who showed conditioning. The results in Table 3 for this response class clearly support the group average results shown in Tables 1 and 2 and again suggest that, with "Good" as the reinforcement, the Humans could be conditioned (and extinguished), whereas the Plurals could not.

These results appear to offer strong support for the conclusion that certain verbal response classes are more amenable to conditioning by use of this type of experimenter verbal reinforcement ("Good") than are other response classes. In addition, the striking similarity between the individual and group results obtained by E_1 and E_2 may be considered to support the conclusion that the phenomena attendant on verbal conditioning are generalizable[5] across at least two *E*s and four different groups of *S*s.

As a further check of these two conclusions, and following upon a suggestion offered to us by Krasner and Sidowski in a personal communication (1959) that the success of our two *E*s with the Human Responses class may have been a function of more experience in verbal conditioning procedures with this class than was the case with our Plural Nouns studies, a fifth study was undertaken by our same two *E*s. In this study, 20 *S*s drawn from the same population were used. E_1 saw 10 *S*s and E_2 10 in an odd-even design. The results for these 20 *S*s (means of 13.0, 11.2, 7.9 for Periods I, II, and III, respectively) again failed to demonstrate conditioning in Period II. Furthermore, analysis by individual *S*s again failed to show (p = .25) a number of conditioners greater than chance expectancy.

To check on the possibility that our two *E*s were themselves somehow contributing to the lack of success with Plural Nouns, a third *E*, described earlier, was used. He had been an *O* in the previous five studies and was thoroughly familiar with the method. Using 10 additional *S*s from the same college population, he obtained Period means for Plurals of 25.8, 26.2, and 28.4, yielding no evidence of conditioning. (We have not yet investigated whether this *E*'s high base rate, 25.8, was due to the smaller N of 10 or whether it represents a stable *E* difference.) Thus, three *E*s and four studies of Plural Nouns failed to demonstrate conditioning for this verbal response class by the use of "Good" or "That's good" as the reinforcement. Because of our inability to condition Plural Nouns and the number of successes with this response class reported in the literature, and because we were able to condition the Human Responses class, we thought it important to investigate a number of variables which usually are not studied but which might shed light on our success with one class and failure with another.

SOME METHODOLOGICAL CONSIDERATIONS

Reliability of E's Scoring

There are many factors which make difficult *E*'s scoring of *S*s' responses in studies such as those reported here. Some of these are: the rapidity with which some *S*s give their serial verbal responses; the low voice level and other verbal characteristics of some *S*s, forcing *E*s to "guess" whether, for example, a word ended with "s" and was thus a plural; rapid judgments by *E* as to whether a word he is hearing for the first time falls in the response class, e.g., "astronaut" (Kanfer, 1958, pp. 443–444, makes this same point); etc. Although we did no systematic study of this variable, we were impressed in a pilot study that one of our *E*s, who "believed" in the phenomenon of verbal conditioning, tended, when in doubt about a word, to score it as "plus" when it fell in Period II, thus enhancing the chance of conditioning. Another of our three *E*s seemed always not to score these doubtful words as positive.

In view of these potential contaminating factors we used as recorders two independent *O*s in addition to *E*. This allowed us to check the reliability of each of these three people against the others. The presence of the other two recorders also reminded *E* to guard against whatever idiosyncratic sets he might have.

For each of the four studies, we were able to compute three reliability coefficients, comparing each recorder against the other two, making 12 in all. Since we computed these reliability coefficients separately for each of the three periods of the experiment, we had 36 separate Pearson *r*'s. Their range (from .96 to .99) indicated an unusually high degree of scorer reliability, thus adding additional confidence to the results here presented.

---

[5] To our knowledge, only one study (Salzinger & Pisoni, 1958) has utilized a second *E* with the *same* group of *S*s as used by E_1. No study has had E_1 retest his own group of *S*s. Test-retest reliability of both types would appear to be very important to our further understanding of verbal conditioning phenomena.

Influence of Characteristics of Individual Ss upon Conditioning

Detailed analyses were carried out on the 80 *S*s who participated in the first four studies. Inspection revealed similar results for the 30 *S*s of the last two experiments. The results for E_1 and E_2 were the same for all variables and thus will be combined.

Age. For the 40 *S*s of the Plurals study, the mean age of the 11 conditioners (22.1 years) was not statistically different from the mean age of the 29 nonconditioners (24.1 years). In the Humans group, the mean age of the 31 conditioners (23.0 years) also did not differ from the mean age of the 9 nonconditioners (25.0). Thus, with an age range from 18 to 47 for the 80 *S*s, there is no evidence that age as a variable influences verbal conditioning or nonconditioning.

Sex. The 80 *S*s included 32 males and 48 females. Of these, 15 males (47%) and 27 females (56%) showed conditioning. This slight difference between the two sexes was not significant. Verbal conditioning or nonconditioning by our two male *E*s seemed independent of the sex of *S*.

Anxiety level. For the Plurals study, the mean Taylor Manifest Anxiety Scale value for the 11 conditioners was 13.18, while for the 29 nonconditioners it was 15.31. In the Humans study, the 31 conditioners had a mean anxiety score of 11.97, while the 9 nonconditioners had a score of 13.67. For neither study, nor for both combined, were these differences significant. Thus, in each of the four separate studies, or when all 80 *S*s were combined, this research failed to show a relationship between Taylor anxiety level and verbal conditioning. This finding is in contrast to that of Taffel (1955), who did find a significant relationship between Taylor MAS score and verbal conditioning. Possibly, the answer regarding the disparate results may lie in the difference between our *S*s (college students) and his (psychotic and neurotic VA patients); or our instructions ("say words") and Taffel's (*S* was asked to make sentences using a verb and one of six pronouns); or our reinforcement ("Good") and his ("Good" for one group, a flashing light for another group, and no reinforcement for a third group); or our method of administration of the Taylor MAS (individual) and his (group).

Vocabulary level. As with the other variables, whether the comparison was of the four studies of 20 *S*s each, or for the 40 *S*s in each of the two response classes, or all 80 *S*s taken together, the mean vocabulary (standard scores) of the conditioners did not differ significantly from that of the nonconditioners. The mean vocabulary score was of the order of 13.0 for each of the four groups. Since the range of scores (8 to 19) for this variable was adequate, verbal conditioning or its absence appears to be independent of the level of *S*'s vocabulary. This may be an important finding since previous research had not dealt with what to us appeared initially to have been an important potential variable in verbal conditioning.

Total time to complete the experiment. Each *E* and the two *O*s kept a record of the total time required by *S* to complete the 400 words over Periods I, II, and III of the verbal conditioning experiment. The range for the 80 *S*s was from 11 to 81 minutes, with an average of 30.72. The mean total time of the conditioners versus nonconditioners did not differ significantly in either the Plurals or the Humans studies. Thus, neither total time nor rate (average number of verbal responses per minute) differentiated either conditioners versus nonconditioners or Plurals versus Humans.

An interesting incidental finding was the nonsignificant correlations obtained in the four studies between Taylor MAS scores and word rate for the whole experiment. The speed with which our 80 *S*s emitted their 400 words was apparently unrelated to their individual anxiety levels. This confirms the negative finding by Benton, Hartman, and Sarason (1955), who studied the word rate in a TAT situation of *S*s scoring high and low on the Taylor MAS.

In summary, with the *S*s used in this study, the results just presented for age, anxiety level, vocabulary level, and total time demonstrate no relevance of these variables to verbal conditioning. However, in each case, the analysis involved a comparison of means (by *F* test) of scores earned by two *discrete* groups (conditioners versus nonconditioners). In this analysis each *S* was weighted equally—he was either a conditioner or a nonconditioner—although his own score earned individual weight within the conditioner-nonconditioner dichotomy. On the chance that such a crude, discrete dichotomization would obscure a real relationship, a score reflecting relative conditioning was obtained by simply subtracting each *S*'s Period I score from his own Period II score.

By use of both product-moment and rho correlational methods, correlations were computed between this individual conditioning score and each of these same four variables (age, anxiety level, vocabulary, and total time). In no case did the correlations reach significance, again suggesting no relationship between verbal conditioning and each of these *S* characteristics. There was a trend for total time to correlate (Pearson *r*) positively with conditioning score for both E_1 (.78) and E_2 (.36) in the Humans study, but comparable correlations were absent (.09 and −.36) in the Plurals group.

On the basis of both the analysis of variance and correlational results, and for the *S*s employed in this study, verbal conditioning, as investigated by the method here described, seems to be independent of such *S* characteristics as age, sex, anxiety level, vocabulary level, and rate of verbal responding.

When Does the Conditioning Period Begin?

Is it appropriate in verbal conditioning studies to begin recording Period II at a time or on a word number predetermined by *E*, or should the beginning of Period II be governed by *S*, individually in each case? More specifically, although *E* might decide that the first 100 words will constitute the baseline or operant period and that the next 200 words will constitute Period II or the conditioning period, what does one do in the analysis for the presence or absence of conditioning in Period II if *S* does not give his *first* response in the category being reinforced until, say, the 52nd word of the 200 words constituting Period II? Obviously, *E* cannot demand that *S* give a Human (or Plural) Response as his first word in Period II so that it can be reinforced according to schedule.

This question is not a new one for investigators in the field of operant conditioning, and they have devised a number of techniques (e.g., "shaping") to speed up the occurrence of the critical response (Skinner, 1953, pp. 91–106; Verplanck, 1955, pp. 598–599). However, since few of the many studies on verbal conditioning reviewed by Krasner (1958) and Salzinger (1959) utilized such shaping techniques, we felt that those investigators who have used the 3-period verbal conditioning methodology employed in the present study had not adequately investigated the pertinence of this question. In general, investigators have merely taken the percentage of critical words in the operant period and have put aside the question of a correction for the ordinal position of the first occurrence of a reinforced word.

Our own empirical results show that while approximately 50% of all *S*s in both the Humans and the Plurals studies gave a word in the appropriate response class—thus obtaining the reinforcement—in the first 10 words of Period II, an additional 30% in the Humans study and 15% in the Plurals study had still not given the appropriate word in the first 30 words of Period II. Even after the first 100 words of our 200-word Period II, 5% (Humans) and 10% (Plurals) of the *S*s still had not received their first reinforcement from *E*. In the Plurals study, 5% of the *S*s still had not received their first reinforcement by the 180th word.

In view of these facts, is it appropriate to begin scoring Period II on a predetermined word number? To do so tends to mask the possibility of a conditioning effect. Accordingly, we analyzed the data for Period II again, separately for E_1 and E_2, for both the Plurals and the Humans classes, this time dividing the number of words given in the response class being reinforced by the number of words (200) in that period minus the ordinal position of the first rewarded word for Period II. For example, if the first reinforcement came on the 40th word in Period II, and *S* then gave a total of 60 words for which he received reinforcement, his Period II score would be 60 divided by 160 or 37.5%. Had one not made this correction, *S*'s Period II score would have been 60 divided by 200 or 30%.

Reanalysis by this method of the data shown in Table 1 changed the mean values for Period II only slightly (21.5% to 22.3%, 15.4 to 15.9, 30.9 to 33.8, and 20.2 to 24.2). No differences across periods for the two Plurals studies emerged, and the significant ($p < .001$) differences for the two Humans studies were maintained.

A similar question could be raised about extinction. Does it begin on a word number predetermined by *E* or on the first appropriate word for which *S* gets no reinforcement, no matter what its ordinal position? Reanalysis of the values for the Period III means shown in Table 1 again yielded corrected values only slightly different from the original ones, and the same findings were obtained: no conditioning for Plural Nouns for either *E*, and conditioning ($p < .001$) of Human Responses by both *E*s.

Verbal Conditioning and Conscious Awareness

The possibility of Thorndikian learning or, more specifically here, verbal conditioning without awareness, has received considerable attention. Excellent discussions of this general area will be found in Postman and Jarrett (1952), Adams (1957), Krasner (1958), and Salzinger (1959). Although many investigators report conditioning without awareness, Adams (1957, p. 391) reports little methodologically sound published evidence to refute the 1952 conclusion by Postman and Jarrett (1952) that "the results [of their own research] . . . show only little reliable evidence for learning without awareness" (pp. 254–255). However, Philbrick and Postman (1955), in a later modification of this earlier study, present evidence that the use of a less complex learning task "facilitated learning without awareness" (p. 424).

In view of the many published positive reports of learning and verbal conditioning without awareness contained in the reviews by Adams (1957), Krasner (1958), and Salzinger (1959), and the questions raised about these studies, especially by Adams, we decided to carry out a detailed analysis of the conscious awareness variable. Although some studies have utilized a post-experiment interview to assess *S*'s level of conscious awareness of the purpose of the experiment, we discarded this interview method after a pilot study revealed that *E*'s behavior during these interviews was too unsystematic. The interview responses of some *S*s forced *E* into a position of almost suggesting the nature of the experiment. In addition, *since the very essence of verbal conditioning experiments is that E can influence S's verbal behavior*, we felt it necessary not to use a method such that *E*'s own biases about whether or not *S*s had attained conscious awareness could in subtle ways influence *S*s' statements as to the purpose of the experiment.

Accordingly, during the post-experiment testing session, we used with each *S* a standardized procedure which employed a single sheet of white paper on which were mimeographed two open-ended statements. At the top of the page was the statement, "The purpose of the experiment was:" In the middle of the page was the statement, "My evidence for this is:" Both statements were followed by a blank space of approximately one half the page, suggesting that *S* could give an answer in considerable detail. The examiner supervised this part of the session very carefully. He took care to see that each *S* understood the two statements. He answered no questions (if any) directly or otherwise suggested answers. Later, by asking *S* (in several cases) to elaborate or otherwise rewrite an answer, he tried to minimize unscorable responses, etc.

It will be clear from the two statements used that the method employed provided information relevant primarily to whether or not *S* understood the purpose of the experiment. We did not question *S* whether or not he was aware of changes in his own verbal behavior. As Adams (1957) points out, these two questions regarding the problem of *S*'s conscious awareness (*S*'s knowing *E*'s purpose versus *S*'s knowing that his own verbal behavior had changed) are quite distinct.

Following the collection of the written conscious awareness statements from the last of the 80 *S*s of our first four studies, each *S*'s degree of "conscious awareness" was determined independently by two judges from *S*'s written responses to the two open-ended statements. The judges were given no information about *S*'s conditioning performance. A four-point rating scale (shown in Table 4) for degree of conscious awareness was developed in a pilot study and was used with the present 80 *S*s. This rating scale is presented in some detail (Columns 1 and 2) in Table 4 because, despite the importance attached to the phenomenon of conscious awareness by so many investigators, published criteria of such ratings have not been available. The illustrative ex-

TABLE 4

CRITERIA FOR LEVEL OF CONSCIOUS AWARENESS RATINGS

Awareness Level	Criteria	Examples	
		S's Statement of Purpose of Experiment	*S*'s Evidence for This Statement
1.	*S* had no idea of the purpose of the experiment, or had a completely wrong hypothesis, or made absolutely no mention of *E*'s (reinforcing) verbalizations.	"To see if I could recall past experiences."	"I realized that the things I was saying were connecting thoughts of wishes, desires, and past experiences which I had had and which surprised me somewhat when I thought how long it had been since I had thought about them."
		"I can't even begin to guess."	None
		"To gather information on whether or not the words people bring up in free-association are of real importance, or of little or no value. Also, it was supposed to reveal a person's speed of communication while under observation."	"The Doctor was keeping a count of my words, and I imagine that the number of words has something to do with an individual's ability to think—fast or slow."
2.	*S* mentioned the reinforcement, but does not connect it with a specific class of words; or *S* brings up the *possibility* of certain words being reinforced (including the correct response class), but in conjunction with other incorrect hypotheses.	"To understand the verbal contrasts between a control group and a group of patients."	"My evidence for this is: (*a*) timing by the instructor for a certain number of words; (*b*) 'good' being said by the instructor on certain words; (*c*) words expressing height, travel, and motivation received 'good'; (*d*) reused words were being counted and investigated to see how often I repeated myself."
		"An experiment on trains of thought. Also perhaps has something to do with amount of *types of words* used."	"It had something to do with trains of thought because to think of a word leads you to think of other words. Therefore everyone would have different thought patterns depending on their physical age and experience maturity. There is a chance that we were to give a definite number of words and maybe the type of word we said was important because, when I said some words he would say 'good.' This made me feel maybe I was saying what he wanted, or what was important, not just ordinary words."
		"To find out how a person relates words and by the relationship understand what areas hold their interest. This doesn't only include academic words, but social words as well. Also, I definitely think the purpose was to see how long your attention can be devoted to one particular object."	"The only evidence I have to back up my statement is when the doctor replied that my words were very good."
3.	*S* stated that *some* specific class of words is being reinforced, but names the wrong response class (e.g., *professions, my family relations*, types of *animals*, etc.); or he states the correct response class (Plurals or Humans) but does so along with a second or third (incorrect) hypothesis.	"To see if a person would continue relating words which he knew were being rewarded by the word good. Then if this reward were suddenly to stop what would be the effect upon the subjects being rewarded for like or the same words."	"Good was said after each word I said to reward that type of word [occupations which serve a purpose in society]. Also several nods were given after the word 'good' was no longer given."
		"I believe the purpose of the experiment was to have the subject name as many different occupations as possible."	"Dr. X would say, 'good,' preceding every occupation that I named, therefore I thought that that was what the purpose of the experiment was."
		"Possibly to see how large ones vocabulary is or to see how many occupational terms one knows."	"After each word such as teacher or doctor the examiner said 'good'."

TABLE 4—*Concluded*

Awareness Level	Criteria	Examples	
		S's Statement of Purpose of Experiment	*S*'s Evidence For This Statement
		"To see what thought channels a person's mind ran in and what subjects were of the most interest to him. The type of words used also helped to reveal level of education in everyday life, also what one would most generally be talking about in everyday life, if not in a test situation."	"On the small common words such as an animal, foods, clothing, etc., there was no facial change on the instructor; however, on peoples' names such as doctors, professional people, Dr. X's expression was more pleasant."
		"To see what effect certain things the interviewer did and said had on the number and kind of words I said. And also the different kinds of words and how they seemed to be selected."	"The interviewer said 'good' after certain kinds of words that I said. I immediately tried to think of more words of this same category or type. I am not sure of which kind but I'll guess: words with a great many different things or other words which would relate to them—possibly *plurals*."
4.	*S* stated that a specific response class (Humans or some very broad and similar category such as people, or Plurals) was being reinforced and *does not* state an alternate, or incorrect hypothesis.	"The experiment seemed to be a relationship of words that had to do with a particular person. The word doctor, geologist, officer, or other words that denoted a title of a person."	"Whenever I mentioned words that pertained to a person I was given a reinforcement of fine or OK from the man in charge."
		"To improve interview procedures. To be able to help lead subjects into a certain line of thought by giving a reward for correct responses."	"Whenever I mentioned words that pertained to people and their occupations I was rewarded with the word 'good,' 'good.' I felt that when I received a response 'good,' that this was the line of thought I should head into."
		"To test peoples reactions to the word 'good.' The word was used in connection with a person or type of person."	"During the middle third of the test the examiner said 'good' or 'That's good' after each word that was associated with a person."
		"To determine whether the subject could recognize what words brought about a 'good' reply from the examiner, and then to see how the subject would react when the words that had once brought about 'good' replies, no longer brought about any reply at all."	"The third word I said was 'schools' which brought about a 'good' reply. When I named further words starting with an 's,' I got no reply. When I named more plural words I got no reply for some time. Then, whenever, I said a plural word, I received a 'good'. All of a sudden this stopped, but I continued to try plural words, after a few unsuccessful attempts at using singular words."

amples shown in Table 4 (Columns 3 and 4) were taken verbatim from the *S*s of the present study.

In order to check the reliability of this four-point scale, the individual ratings of the two judges were compared. The Pearson *r* for their independent judgments of the conscious awareness levels of the 40 *S*s in the Plurals study was .92; for the 40 *S*s of the Humans study, *r* was .85. The corresponding values for rho were .92 and .87. These coefficients (at the .001 level of confidence) indicated a sufficiently high degree of reliability for the rating scale shown in Table 4 to permit analysis of the relationships between each *S*'s expressed level of conscious awareness and his conditioning performance.

First to be examined were group trends for the degree to which *S*s in the two response classes were able to verbalize the purpose of the experiment. The results shown in Table 5 offer clear evidence that, as a group, the 40 *S*s in the two Humans studies were significantly ($p < .001$) more aware of the relationship between their own verbal behavior and *E*'s than were the 40 *S*s in the two Plurals studies. The difference between the awareness level associated with each of the two response classes is approximately one whole unit toward the full conscious awareness end on the four-point

TABLE 5

MEAN RATINGS OF SUBJECTS' CONSCIOUS AWARENESS LEVELS FOR TWO RESPONSE CLASSES

Judge		Plural Nouns ($N = 40$)	Humans ($N = 40$)	F
1	M	1.58	2.50	19.22***
	SD	.86	1.00	
	R	1–4	1–4	
2	M	1.55	2.25	12.10***
	SD	.86	.92	
	R	1–4	1–4	
Both	M	1.56	2.38	15.35***
	SD	.84	.92	
	R	1–4	1–4	

*** $p < .001$.

awareness scale (mean ratings of 1.56 versus 2.38 when the ratings of two judges were combined). Thus, the group data clearly suggest that the verbal conditioning obtained in the Humans studies here reported (Table 1) was associated with *S*'s greater ability to verbalize (or have conscious awareness of) the purpose of *E*'s reinforcing behavior. In many cases, the written responses clearly indicated *S*'s awareness of the contingency between *E*'s behavior and changes in his own verbal responses.

Additional evidence for this point came from a further breakdown (not shown in Table 5) of the group data. The 40 *S*s of both the Plurals and the Humans studies were again divided into conditioners versus nonconditioners, and the mean conscious awareness scores were computed separately for each response class. The results for the Humans study are that the 9 *S*s who did not show conditioning reported a significantly ($p < .01$) lower level of conscious awareness (1.56) than did the 31 *S*s who did show conditioning (2.55). For the Plurals study, while not statistically significant, the 11 conditioners showed a trend toward more awareness (1.91) than did the 29 nonconditioners (1.45). On the basis of this more stringent analysis of the group data, it can be concluded, especially in the two Humans studies, that conscious awareness as here defined was a concomitant of verbal conditioning. A similar (nonsignificant) trend held for the Plurals group data, although it should be remembered that for this response class no statistically demonstrable conditioning occurred.

In order to further evaluate the group results for these two response classes, we again investigated the performance of individuals within the groups. It was our hypothesis that if degree of conscious awareness was associated with absence or presence of verbal conditioning, then one should find more *S*s showing conditioning than not showing it as level of conscious awareness increased from Level 1 through Level 4.

The results shown in the bottom half of Figure 3 are a clear demonstration of this hypothesis for the 40 *S*s in the Humans study. Whereas one would expect by chance alone one *S* out of four to show "conditioning" at each of the four awareness levels, this chance occurrence is seen only for the 8 *S*s who reported an awareness level of 1. That is, for this stringent "no awareness" level, the 3 *S*s out of 8 (37.5%) who showed "conditioning" are not significantly different from the 2 out of 8 (25%)

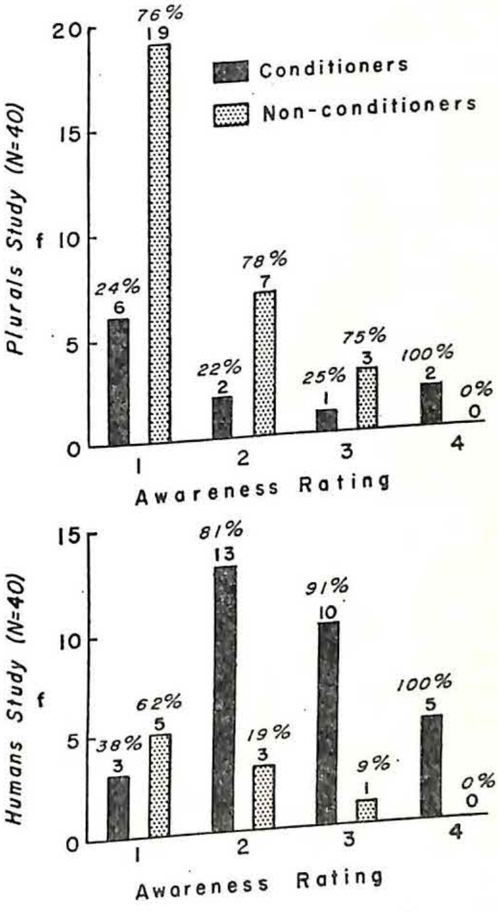

FIG. 3. Frequency distribution at each awareness level.

expected by chance alone to show an increase in Period II relative to Period I. This is not true for each of the three higher awareness levels (Levels 2, 3, and 4). For each of these three levels, the data for Humans strikingly demonstrate the fact that the number of *S*s who conditioned far exceeded the one in four (25%) expected by chance alone (i.e., 81.2, 90.9, and 100% of the *S*s showed conditioning at Awareness Levels 2, 3, and 4, respectively). Furthermore, despite the small *N*s involved, it is clear from these figures that, as level of conscious awareness increased from Level 2 through 4, the percentage of *S*s showing conditioning also increased in stepwise fashion. Thus, from a chance relationship at Level 1 (37.5%), the percentage increased to 81.2% (13 out of 16) at Level 2, 90.9% (10 out of 11) at Level 3, and 100% (5 out of 5) at Level 4.

As shown in the upper half of Figure 3, the results of a similar analysis by individuals for the 40 *S*s in the Plurals studies again failed to reveal any evidence of a relationship between conditioning and conscious awareness for this response class. Except for Level 4, where both *S*s reporting this degree of awareness showed conditioning, the number of *S*s "conditioning" at each of the three other levels was no greater than the 25% who would be expected by chance alone (i.e., 24, 22.2, and 25%, for Levels 1, 2, and 3, respectively).

In order to provide a statistical analysis of the data shown in Figure 3, a chi square was computed individually for each of the two response classes. This consisted of a chi square comparison of conditioners versus nonconditioners for the *S*s with an awareness rating of 1 (minimal or no awareness) and a similar single chi square comparison for *S*s earning all the higher awareness ratings taken as a group. For the Plurals data, the value of chi square did not begin to approach an acceptable level of significance either for the 25 *S*s in Group 1 (chi square of .00), or for the 15 *S*s in the combined Groups 2, 3, and 4 (chi square of .55). For the Humans data, the number of conditioners versus number of nonconditioners for the 8 minimal awareness *S*s was equally nonsignificant (.66); whereas for the remaining 32 *S*s it was highly significant (66.67, $p < .001$). Thus, both the analysis by groups (Table 5) and the analysis by individuals (Figure 3) demonstrate that, in the studies here reported, verbal conditioning or its absence was quite clearly associated with *S*'s degree of reported conscious awareness. It is possible, therefore, that one of the reasons for our failure to condition Plural Nouns, while other investigators have succeeded, lies in the differences in awareness levels of our *S*s relative to the *S*s (and other experimental conditions) employed in these other studies.

TABLE 6

RELATIONSHIP BETWEEN CONDITIONING SCORE AND AWARENESS LEVEL

	N	*r*	*r* with Transformed Scores
		PLURAL NOUNS	
E$_1$	20	.64***	.54**
E$_2$	20	.31	.25
Both *E*s	40	.48***	.40***
		HUMANS	
E$_1$	20	.38	.41
E$_2$	20	.49*	.55**
Both *E*s	40	.45***	.50***

* $p < .05$.
** $p < .02$.
*** $p < .01$.

The fact that the verbal conditioning of our *S*s was correlated with *S*'s level of conscious awareness was demonstrated in still another way. As previously mentioned, a dichotomy into conditioner versus nonconditioner, such as is presented in Figure 3, gives each *S* who shows conditioning equal weight to every other *S* who shows conditioning, regardless of magnitude. In order to allow for each *S*'s actual conditioning performance, the previously described conditioning score for each *S* (the difference between *S*'s Period II minus Period I score) was correlated with his reported conscious awareness score. In view of the high correlations between the two judges for this awareness score, the two judgments were pooled for an average awareness score.

The results of this analysis are shown in Table 6. Due to the restricted range (1 to 4) of the awareness ratings, as well as the presence of skewed distributions of scores for both variables, a simple square root transformation (Snedecor, 1946, pp. 445–447) was made. These results are shown along with the Pearson *r* values in Table 6. Values for rho also were computed and yield similar results.

The data presented show a moderate but significant correlation between magnitude of conditioning score (increase from Period I to Period II) and level of later reported conscious awareness. The positive correlations and an inspection of the frequency distribution of conditioning scores indicate that the greater the increase from Period I to Period II, the higher was *S*'s later reported level of conscious awareness. Three of the 40 *S*s in the Human Responses class showed a decrease in Period II relative to Period I, whereas 20 of the 40 *S*s in the Plurals study showed such a decrease. Fourteen of these 20 *S*s fell in awareness Level 1. For Levels 2, 3, and 4, the frequency distribution for Plurals (i.e., the conditioning score), like that for Humans across Levels 1, 2, 3, and 4, showed a positive correlation between amount of conditioning and degree of later reported conscious awareness. This finding of a significant positive correlation between magnitude of conditioning score and level of awareness for the Human Responses is not surprising in view of the results shown in Figure 3. That the dichotomous analysis shown in Figure 3, as well as the chi square analysis, did not demonstrate such a similar relationship for the Plural Nouns class, whereas the correlational analysis shown in Table 6 for this same response class did, would suggest that magnitude as a measure of conditioning related to level of conscious awareness may be more sensitive than dichotomous classification. In any event, all analyses demonstrate the same finding: for our *S*s, verbal conditioning occurred in those *S*s who later could verbalize the purpose of the experiment. This was especially true for the Human Responses class (Figure 3 and Tables 5 and 6), but also true for one analysis of the Plural Nouns data (Table 6). The minor differences shown in Table 6 between the results obtained by the two *E*s appear to be nonsystematic and, thus, probably are due to random sampling errors.

Influence of S's Base Rate on Verbal Conditioning

In view of the significant differences in the base rates of the Humans versus Plurals response classes (Table 2), it appeared wise to investigate further the role of these base rates in conditioning, especially as base rates and conditioning might in turn be related to *S*'s level of conscious awareness. Salzinger and Pisoni (1958, p. 88) reported a finding which bears indirectly on this question. They reported a cross-validated rank-order correlation of .46 between the number of interview Affect responses given by 20 schizophrenic *S*s in a 10-minute base period and the number of Affect responses given in the next 10-minute (conditioning) period. Sixteen additional *S*s, serving as a control group, and with whom *E* attempted no conditioning, showed a still higher correlation (rho of .85, $p < .01$) between the first and second 10-minute periods. Thus, although smaller in magnitude (rho of .46), there still appeared to be a correlation between base period and conditioning period levels for this (Affect) response class. However, it also is clear from these rho values that when *E* applied his verbal reinforcement ("mmm-hmm," "I see," "yeah," and other forms of agreement) and thereby produced conditioning in the experimental group, the correlation between base rate level and Period II level dropped from .85 (control group) to .46 (experimental group).

Similar rank-order correlations between Period I and Period II are shown in Table 7. It is clear from these results that, for both response classes, there is a significant positive correlation between the percentage of words given in the base period and the percentage of similar words given in the conditioning period. Since we failed to obtain conditioning for the Plurals response class, the 40 *S*s in this response class may be thought of as a control group crudely analogous to the one used by Salzinger and Pisoni. Such a conception would

TABLE 7

RELATIONSHIP BETWEEN SUBJECTS' PERIOD II AND BASE RATES

	N	Plural Nouns rho	Humans rho
E_1	20	.68**	.20
E_2	20	.56**	.47*
Both *E*s	40	.61**	.36*

* $p < .05$.
** $p < .01$.

offer a tentative hypothesis for explaining the somewhat smaller, albeit still significant, rhos obtained by us for the Human Responses class. That is, it is possible to conclude that the correlation between *S*'s verbal behavior in Period I and Period II may be positive but small when *S* is aware of the contingency between his own behavior and *E*'s (our Humans study), but may be much greater either when *E* is applying no reinforcement (Salzinger and Pisoni), or when *S* is unaware of the reinforcement (our Plurals study), or when reinforcement is having no effect for other reasons.

In view of this demonstrated small positive correlation between base rate and Period II response level, especially in the Humans study, we decided next to investigate the relationship between base rate and the earlier described magnitude of conditioning score, i.e., how much an *S* increased (or decreased) his score from Period I to Period II. We reasoned that this conditioning score might turn out to be independent of *S*'s base level because it represents the *difference* between Period I and Period II scores.

Because of the presence of skewed distributions, both rhos and Pearson *r*'s were used for computing the correlation between each *S*'s base rate and his subsequent conditioning (difference) score. For the Human Responses class, and for each *E* and for both *E*s combined, the two types of correlations between base rate and conditioning score were found to be near the zero level. Thus, conditioning of the Human Responses class, as demonstrated in Table 1 and in the analysis by individuals (Table 3), appears independent of each *S*'s own operant level.

Interestingly, the same lack of relationship was not true for the Plurals response class. For this category, a small negative correlation ($-.37$, $p < .02$) was obtained between base rate for Plurals and conditioning score. The *S*s of both *E*s obtained these small negative correlations, indicating that the higher the base rate, the smaller was the increase from Period I to Period II and vice versa.

One possible explanation for this finding of a small negative correlation in the Plurals studies but not the Humans studies is that the negative correlation between base rate and conditioning score is unrelated to the magnitude of the conditioning score but, rather, is related to the magnitude of the base rate. This hypothesis is based on the demonstrated ($p < .001$) difference in base rates for the two response classes (Table 2)—16% for Plurals versus 6% for Humans—and the correlated fact that an *S* with a high base rate (as in the two Plurals studies) did not have the same opportunity to increase his score in Period II to the same degree as an *S* with a lower base rate (as in the two Humans studies). Thus, the *S* with a large base rate score would have a higher probability of earning a lower conditioning score, thus yielding a negative correlation between base rate and conditioning score, than would the *S* with a smaller base rate score. Until further research bearing on this hypothesis is completed, we believe it to be an adequate one to explain our findings.

Thus the results of our own study (Table 7) and Salzinger and Pisoni's (1958) suggest a significant positive correlation between actual scores in Period I and actual scores in Period II. The small significant negative correlation (between Period I score and conditioning score) just described indicates a qualification to this inference. Whereas an *S*'s Period II score can be predicted, within limits, from his own Period I or base level score (Table 7), the magnitude of his change from Period I to Period II is inversely related to the magnitude of his base level score for one response class (Plurals) and has no relationship for another response class (Humans), presumably because this latter response class has a lower operant or base rate. Furthermore, the negative correlations obtained for one response class (Plurals) and not for another (Humans) again demonstrate the necessity for careful study of different verbal response classes (including their free-operant levels) before the results of verbal conditioning studies are generalized too far.

Why Are Some Verbal Response Classes Easier to Condition than Are Others?

At this stage of our knowledge no adequate data exist to answer this question of differential effects over response classes. Although the results of our own studies show that *S*'s degree of conscious awareness is correlated with whether or not he shows conditioning, the fact still remains that we were able to produce both this higher level of awareness and a change in *S*'s behavior in Period II ("conditioning") with the Human Responses category but not with the Plural Nouns category. Thus, the two response classes must differ along at least one crucial dimension. Table 2 shows that the free-operant frequency of emission of these two classes of responses differs significantly. It is our own belief that the two response classes differ in still another way, in their level of difficulty as concepts. That is, considered purely in terms of difficulty of concept formation, it is reasonable to predict that *S*s could more readily abstract or perceive the similarity among five verbal responses like mother, neighbor, friend, architect, plumber than among books, apples, cars, shoes, flowers. If this is true, then Humans as a response class may be more easily conditioned than the more difficult concept Plurals. This would be especially true if *S*'s level of conscious awareness is related, as is the case here, to his conditioning performance. In any event, the difficulty level of the concept being conditioned appears to be a variable worth exploring. This variable may turn out to be as important as the others already identified (reinforcement used, base rates, *S*'s level of reported conscious awareness, etc.).

Another possible explanation for our success in conditioning Humans and not Plurals has been offered by B. F. Skinner (personal correspondence, 1960). It involves his concepts of the discriminative stimulus, S^D, and the prior reinforcement and conditioning history of the individual. His suggestion runs as follows:

> It seems to me quite reasonable that the so-called reinforcement in period two acts more as a discriminative stimulus than as a reinforcement. In talking to the lady on my right at a dinner party, I may try various subjects without much success. When I hit on something which gets a nod of interest, I continue along the same line, not so much because the response has been reinforced, as because the response plus the reinforcement act as an S^D indicating that this is the kind of audience which reinforces certain kinds of responses. The class of responses has already been learned in the past. The first "Good" in period two may be closer to a green light which serves as a discriminative stimulus for "more of the same." If this is the case, you could probably show that the class of *humans* has already been brought under audience control where the class of *plurals* has not. Thus, one could discover that a given listener is interested in *people*, although it would be quite rare to discover that a given listener is interested in *plurals*. This would throw the explanation back on the prior conditioning of the subject with respect to audience variables....

With notable exceptions (Salzinger, 1959, pp. 85–87), investigators working in the area of verbal conditioning have tended to ascribe the conditioning effect to the reinforcement used by *E*, rather than to think in terms of S^D and the individual's prior differential conditioning history to disparate verbal response classes.

Another variable which may have been responsible for our success with the Human Responses class and failure with Plurals may be an "expectancy-set" induced in our *S*s by the physical setting in which the experiments were conducted. Since the *S*s of this study were studied in the laboratories of the psychology-psychiatry service of a large medical center, the *S*s, in line with cultural conditioning regarding psychiatry, may have been "set" to talk about themselves and other people, thus showing a conditioning effect to this "expected" Human Responses class and not to the "not-expected" Plurals class. Thus "situation-induced expectancy" as a set-variable may be pertinent in understanding our results. This is especially so since our *E*s wore their usual white hospital coats during the conditioning experiments, and thus may have enhanced the set of seeking responses of an interpersonal (Humans) nature in our *S*s.

SUMMARY

Two *E*s, in replication experiments, independently studied two groups of 20 college undergraduates each, making a total of 80 *S*s in all. As additional checks on the results, the same two *E*s studied a fifth group of 20 *S*s, and a third *E* studied a sixth group of 10 *S*s. Verbal conditioning was attempted in separate studies with two verbal response classes: Plural Nouns and Human Responses. Directions called essentially for *S* to "say words, but not sentences or numbers." The *E* used three periods of 100, 200, and 100 words, constituting the operant, conditioning, and extinction periods, respectively. The statement "Good" or its variant, "That's good," was used by *E* as the reinforcement in Period II.

Results indicate that neither *E* could condition Plural Nouns, whereas both *E*s successfully conditioned the Human Responses class. Thus, response class is an important variable

in verbal conditioning. Since a number of different analyses all revealed that for both *E*s the successful conditioning of the Human Responses class was associated by *S*s in these groups with a higher level of conscious awareness of the purpose of the experiment than was the case with *S*s in the Plurals groups, it was demonstrated that the presence of "conscious awareness" was a concomitant of conditioning. Level of *S*s' conscious awareness was determined independently by two judges from written statements to two open-ended questions obtained from *S* immediately after the verbal conditioning experience. The four-point rating scale used for these judgments was shown to have high interjudge reliability. Further analysis revealed that under the conditions of these studies, verbal conditioning was independent of such characteristics as age, sex, vocabulary level, (Taylor) anxiety level, and word rate.

In order better to understand the phenomenon of verbal conditioning, a number of additional methodological considerations were discussed: reliability of *E*'s scoring, determination of when the conditioning period begins, influence of *S*'s base rate on verbal conditioning, and influence on verbal conditioning of the level of difficulty of different response classes as concepts. Analyses of the data of the present studies were used in the presentation of these methodological considerations.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

BENTON, A. L., HARTMAN, C. H., & SARASON, I. G. Some relations between speech behavior and anxiety level. *J. abnorm. soc. Psychol.*, 1955, **51**, 295–297.

EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

KANFER, F. H. Verbal conditioning: Reinforcement schedules and experimenter influence. *Psychol. Rep.*, 1958, **4**, 443–452.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.

MANDLER, G., & KAPLAN, W. K. Subjective evaluation and reinforcing effect of a verbal stimulus. *Science*, 1956, **124**, 582–583.

PHILBRICK, E. B., & POSTMAN, L. A further analysis of "learning without awareness." *Amer. J. Psychol.*, 1955, **68**, 417–424.

POSTMAN, L., & JARRETT, R. F. An experimental analysis of "learning without awareness." *Amer. J. Psychol.*, 1952, **65**, 244–255.

SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, **61**, 65–94.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, **57**, 84–90.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Iowa State College Press, 1946.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

VERPLANCK, W. S. The operant, from rat to man: An introduction to some recent experiments on human behavior. *Trans. NY Acad. Sci.*, 1955, **17**, 594–601.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychol. Corp., 1955.

WILSON, W. C., & VERPLANCK, W. S. Some observations on the reinforcement of verbal operants. *Amer. J. Psychol.*, 1956, **69**, 448–451.

(Early publication received June 16, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 207-215

ANXIETY IN ACADEMIC ACHIEVEMENT SITUATIONS[1]

RICHARD ALPERT
Harvard University

AND

RALPH NORMAN HABER
Yale University

As part of the effort to define and measure the variables of which test performance is a function, this study was designed to evaluate, both experimentally and theoretically, the paper-and-pencil instruments currently being used in American research for the measurement of individual differences in anxiety as it affects academic achievement performance. In addition, it includes a description of a new achievement-anxiety scale which has been devised to indicate not only the presence or absence of anxiety, but also whether the anxiety facilitates or debilitates test performance.

Three separate problems are considered: (*a*) the relationship between scales which are designed to measure general anxiety and scales specifically designed to measure test anxiety (specific anxiety scales), and a comparison of the relative efficacy of the general and specific scales as predictors of academic achievement performance; (b) the relationship between the construct of anxiety and that of aptitude and the methodological problems involved in separating these two operationally; and (*c*) the direction of the effect of anxiety upon academic achievement performance.

Procedure and Subjects

Procedurally, this research involved the administration of a variety of anxiety scales to introductory psychology students and to freshmen at Stanford University. The scales used included the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953), the Welsh Anxiety Index (AI) (Welsh, 1952), the Freeman Anxiety Scale (AS) (Freeman, 1953), the Mandler-Sarason Test Anxiety Scale (TAS) (Mandler & Sarason, 1952), and the Achievement Anxiety Test (AAT), which has two scales discussed below. Analysis of the data involved intercorrelating these scales and studying their relationship to a measure of verbal aptitude, the Scholastic Aptitude Test (SAT) (College Entrance Examination Board, 1956), and to a set of academic performance indices, including college grade-point average and the final examination, mid-term examination, and course grades in the introductory psychology course. Only male *S*s were included in the data analyses. For the students enrolled, the scales were either administered during the first class meeting of the quarter to the entire class (Fall 1955, Winter 1956, and Spring, 1956, with *N*s of 93, 92, and 98, respectively), or as part of smaller group experiments which were required of all students (Spring, 1956, subsample, $N = 40$). Therefore, these data are representative of the population of the students from the introductory course in psychology, which is taken by 75% of the students in the university. For the freshmen, the scales were group administered as part of the freshman evaluation program.

The scores for the Scholastic Aptitude Test (administered to the students during their senior year in high school) and the grade-point averages were obtained from the Registrar of the university. The final examination, course, and midterm grades were obtained from class records. All of these data were collected without the *S*s' awareness either that it had been collected or that data of this nature would be related in any way to their responses on the anxiety scales.

In view of the finding of Davids (1955) that scores on anxiety scales are affected by the *S*s' particular motivation for filling out anxiety questionnaires, it must be noted that *S*s completed the present questionnaires with the understanding that the data were being collected as part of a large ongoing research program conducted by the Psychology Department and that their answers would be used for research purposes only. In some instances, *S*s' affiliation motivation might have been aroused when, for example, the entire class was asked to fill out a questionnaire as a favor to the experimenter; at times when the questionnaire was administered in experimental session, the fulfillment of course requirements was the major motivation. In all situations, however, it was emphasized that the data would *not* be made available either to the university administration or to the faculty and that no selection of any kind would be based on the results of the questionnaire. There is no reason to believe that motivational differences in any way differentially affected the responses.

General versus Specific Anxiety Scales

The first concern in constructing a measuring instrument which will predict the effect of anxiety in academic performance is one of defining the population of experiences and behaviors which the item responses are to reveal. There are current in the literature two divergent positions regarding this matter. One position is exemplified by Taylor (1953) with her Manifest Anxiety Scale and the other by Mandler and Sarason (1952) with their Test Anxiety Scale.

---

[1] The support provided by the National Science Foundation (G-1787 and G-3045) is gratefully acknowledged. We also wish to thank R. R. Sears, Q. McNemar, and Jan Pierce for their help and support, and the Stanford University Office of the Registrar and Computing Center for their cooperation.

Part of this report appeared as a portion of a doctoral dissertation at Stanford University by the first author.

Taylor's general anxiety scale is made up of items drawn from the MMPI and is concerned with a wide variety of situations other than test taking. Underlying the construction of the MAS is a theoretical assumption that there is a relatively constant "level of internal anxiety or emotionality," and also "that the intensity of this anxiety could be ascertained by a paper-and-pencil test consisting of items describing what have been called overt or manifest symptoms of this state" (Taylor, 1953, p. 285). If Taylor is correct in positing such a general anxiety state, then it should follow that a single measure of a set of manifest anxiety responses gathered from many situations would be an adequate predictor of the presence and effects of anxiety responses in *any* situation, whether it be eyelid conditioning (Spence & Taylor, 1951; Taylor, 1951), therapeutic sessions (Peck, 1951), serial learning (Taylor & Spence, 1952), or academic performance. At present, however, there is some evidence (Child, 1954; Mandler, 1954) of *intra*-individual differences in anxiety both in content and intensity from one situation to another. This evidence warrants further examination of the basic general-anxiety-state assumption underlying the construction of the Taylor MAS-type scale before applying such a scale as extensively as has been done in the past. Too often, conclusions have been drawn on the basis of correlations with the MAS regarding the presence or absence of anxiety or the effects attributable to anxiety without due consideration of the possible limitations of a general scale of this type as a sensitive indicator of anxiety in any limited, recurring type of situation.

Mandler and Sarason, who represent the alternative current position, maintain that the items composing the measuring instrument should be concerned with the specific situations in which it is to be used. These investigators, studying stress in academic achievement situations, devised the Test Anxiety Scale (TAS), ". . . a questionnaire which was specifically concerned with the *S*s attitudes and experiences in a testing situation" (Mandler & Sarason, 1952, p. 166). Thus, the TAS provides a score indicating the recalled intensity of certain experiences and behaviors immediately antecedent to or concomitant with the taking of various types of examinations in the past. The implication of this type of scale is that the increased situational specificity of its item content will allow for a more sensitive measurement of anxiety and its effect in the academic achievement situation. It follows that for other situations other specific scales will be needed; and, only if these specific scales, each shown to be valid in its own situation, turn out to be highly intercorrelated may one posit a general anxiety state and thus justify a single measure of anxiety for all situations.

The relative merits of situational specificity versus generality of item content can be evaluated by two statistical techniques. The first is a comparison of the intercorrelations among a variety of general and specific scales administered to the same subjects. The Taylor MAS, the Welsh AI, and the Freeman AS are used as measures of general or manifest anxiety; the Mandler-Sarason TAS and the two scales of the AAT are used as measures of specific anxiety. High intercorrelations among all six scales would support the theoretical assumption of a single underlying state, and thus the appropriateness of a single general measure. On the other hand, low intercorrelations would support the use of scales which are specific to the situations in which they are to predict. Table 1 reports the intercorrelations of the measures of anxiety for an *N* of 40 (the only sample that received all six scales). Because in all instances inspection of the scatter diagrams revealed no departure from linearity, only product-moment correlation coefficients were computed.

As Table 1 indicates, the correlations among the general anxiety scales range from .32 to .39. The correlations between the general and specific scales range from .24 to .38. The correlations among the specific anxiety scales range from .40 to .64, all of which appear to be higher than any of the correlations involving the general scales. This observation is supported by significance tests of the differences between correlations, which indicate that in most instances the intercorrelations among

TABLE 1

INTERCORRELATION AMONG SIX MEASURES OF ANXIETY ($N = 40$)

Test	REL	AI	AS	TAS	AAT−	AAT+
General Anxiety Scales:						
MAS	.89	.39*[a]	.32*	.32*	.38*	−.33*
AI	.84		.34*	.28	.37*	−.25
AS	.73			.38*	.30	−.24
Specific Anxiety Scales:						
TAS	.82				.64**	−.40*
AAT−	.87					−.48**
AAT+	.83					

Note.—Reliabilities of the scales appear in Column 1.

[a] Based on the nonoverlapping items. When overlapping items are included, $r = .89$.

* Significant at .05 level.

** Significant at .01 level.

TABLE 2

CORRELATIONS BETWEEN SIX ANXIETY SCALES AND FOUR MEASURES OF ACADEMIC PERFORMANCE ($N = 40$ to 379)

Test	Grade-Point Average	Course Grade	Final Exam Grade	Midterm Exam Grade
		Psych 1	Psych 1	Psych 1
General Anxiety Scales:				
MAS	.01	−.08	−.02	−.19
AI	−.04	−.05	−.03	−.22*
AS	−.06	.14	.15	—
Specific Anxiety Scales:				
TAS	−.24*	−.21	−.16	−.32*
AAT−	−.35*	−.26*	−.28*	−.25*
AAT+	.37*	.23*	.26*	.21

* Significant at .05 level.

the specific scales are significantly higher than the correlations involving the general scales either with each other or with the specific scales. This seems to throw some doubt on the comparability and, therefore, substitutability of a general anxiety scale for a specific anxiety scale.

A factor analysis was also performed (Alpert, 1957) on this matrix of correlations, but although the above conclusions were supported by this analysis, such support is little more than suggestive because of the hazards of drawing conclusions from a factor analysis based on only 40 cases.

A second technique for determining the substitutability of the general for the situationally specific anxiety scales, far more rigorous than merely intercorrelating the scales, is to compare the two types of scales as predictors of various measures of academic performance. If the general scales predict as well as the specific scales, there is justification for the use of a single general measure for the detection of anxiety in all situations. If, on the other hand, the specific scales are able to account for significantly more of the variance in the performance measures than are the general scales, then there is justification for the use of scales which are specific to the situation in which they are used.

Table 2 presents the correlations between the six anxiety scales and four measures of academic performance. These are average correlations, based on combined samples with total *N*s ranging from 40 to 379. Thus, these relationships are based on all samples for which both the appropriate anxiety and performance scores were available. Before averaging was attempted, a check was made to be sure that neither the means and the standard deviations of all the samples nor the correlations to be averaged were significantly different. No differences were found in any of these comparisons.

The data summarized in Table 2 indicate that the specific anxiety scales are more often significantly correlated with academic performance measures than are the general anxiety scales. Accepting the 5% level of significance (two-tailed) as a criterion, only one of the 11 available correlations for the general scales reached significance, while 9 of the 12 correlations involving the specific scales reached significance.

The implications of the findings are reasonably clear. Specific anxiety scales and general anxiety scales measure, to a significant extent, something different. Furthermore, it appears that the variable which the specific scales measure, and which the general scales do not, is involved in academic performance to such an extent that the specific scales are better predictors of academic performance than are the general anxiety scales.

One may tentatively conclude, therefore, that further use of a general anxiety scale as

TABLE 3

CORRELATIONS BETWEEN SIX ANXIETY SCALES AND A MEASURE OF VERBAL APTITUDE (SAT)
(N = 40 to 379)

Test	Verbal Aptitude		
	N	r	p
General Anxiety Scales:			
MAS	198	+.10	*ns*
AI	153	+.13	*ns*
AS	40	−.24	*ns*
Specific Anxiety Scales:			
TAS	154	−.18	.05
AAT−	379	−.29	.001
AAT+	379	+.21	.001

an appropriate operation for the measurement of academic achievement anxiety is unwarranted. However, the next section, on the relationship of these various anxiety scales to aptitude must be considered before this can become a final conclusion.

THE RELATIONSHIP BETWEEN ANXIETY AND APTITUDE

In evaluating the efficacy of the various scales in the academic situation, one must take cognizance of the relationship between the anxiety measures and intellectual ability because interest is centered in a scale which predicts performance variance attributable to something other than aptitude. On this basis, it seems desirable for an anxiety scale to correlate highly with performance and only slightly with aptitude.

The equivocal results reported in the literature regarding the relationship between anxiety and aptitude become more clear if one takes into consideration the type of anxiety scale used, the heterogeneity of the population measured, and the nature of the aptitude measure (most importantly, whether it was timed and, therefore, under pressure or untimed). (See Alpert, 1957, for a more extensive review of these variables.) Since the present study is concerned with a highly intellectually homogeneous college population and with an aptitude test given under pressure conditions, the literature involving comparable populations and test conditions is of most direct concern. It indicates that the general anxiety scales are not related to timed aptitude tests in homogeneous college populations (Davids & Eriksen, 1955; Schulz & Calvin, 1955), whereas specific anxiety scales are related to this kind of aptitude test (Sarason & Mandler, 1952). The data presented in Table 3 support these findings. Here the general anxiety scales are not significantly related to aptitude, but all the specific scales are significant beyond at least the 5% level.

There are four possible explanations for this difference:

1. The specific scales are in part a measure of intellectual ability. That is, for some reason, intelligence, *independent of actual anxiety level*, affects the individual's response to a specific anxiety measure. General scales are free of an intellectual ability component according to this argument.

2. A specific anxiety scale is an appropriate vehicle for rationalization or justification of poor academic performance. Because people of lower intelligence are more apt to have experienced the effects of the results of poor academic performance in the past, more of them would be inclined to use such rationalization than would people of higher intelligence, and the specific scales are more sensitive to rationalization attempts.

3. There are stress cues connected with the taking of timed college aptitude tests. These cues elicit anxiety responses which affect the aptitude test performance. The specific scales are sensitive to the presence of anxiety of the type reflected in the aptitude tests, but the general scales are not similarily sensitive.

4. The more intelligent the individual, the less anxiety he manifests in test situations because he has less objective reason to fear the experience. The specific scales, which are able to measure this interaction, are, therefore, a sensitive index of anxiety in a college population, whereas the general scales are not.

The last of these arguments is the only one which maintains that the obtained correlation is a measure of the relationship between the two constructs of anxiety and aptitude. The first three suggest that the correlation between the measures is an artifact which appears because of poor test construction or the conditions of measurement. While the present studies were not designed to evaluate these various explanations, they do provide suggestive evidence of their relative validity.

The first explanation holds that if an anxiety scale which is correlated with an aptitude measure is, because of some characteristic such as its format, simply a measure of aptitude itself, then one would expect that the aptitude and anxiety measures would account for the same variance in academic performance. If this were, in fact, the case, then a multiple correlation based on the best weighted combination of aptitude and anxiety would not be expected to account for significantly more of the variance in academic performance than would aptitude alone. Tables 4 and 5 present multiple correlations of aptitude and anxiety, predicting grade-point average and final examination grade. Also, they indicate the significance of the addition of the anxiety measure in the predictions. The general anxiety scales are not included in the tables because they are not correlated with achievement performance and hence, could make no contribution to the prediction of such performance.

The debilitating scale of the AAT adds significantly, over and above aptitude, to the prediction of the grade-point average (GPA) in two out of three instances, and in one out of three for the final examination grade. The facilitating scale of the AAT adds significantly, over and above aptitude, to the prediction of GPA in the three instances reported, and in two out of the three instances for the final examination grade. The Mandler-Sarason TAS adds significantly to prediction of final examination grade, but not for GPA, in the only samples which data are available. These findings refute the argument that the specific scales which are correlated with aptitude are simply measures of aptitude, because the results indicate that specific anxiety scales are able to account for variance in academic performance other than that accounted for by a measure of aptitude. This does not mean to say that these anxiety instruments are not, in part, measures of intellectual ability; it does say, however, that they are something more than that—and that therefore they are able to account for added variance in academic performance, and, as such, are valuable supplements to a measure of aptitude.

TABLE 4

MULTIPLE CORRELATIONS PREDICTING GRADE-POINT AVERAGE (GPA) FROM A COMBINATION OF VERBAL APTITUDE AND ANXIETY

Test	N	R^a	r^b	r^c	p^d
AAT−	93	.46	−.45	.27	.001
AAT−	92	.29	−.08	.29	*ns*
AAT−	98	.53	−.40	.43	.001
TAS	40	.46	−.27	.41	*ns*
AAT+	93	.42	.36	.27	.01
AAT+	92	.43	.32	.29	.01
AAT+	98	.58	.50	.43	.001

[a] GPA · Aptitude & Anxiety.
[b] GPA · Anxiety.
[c] GPA · Aptitude.
[d] For adding anxiety to aptitude.

TABLE 5

MULTIPLE CORRELATIONS PREDICTING FINAL EXAMINATION GRADE FROM A COMBINATION OF APTITUDE AND ANXIETY

Test	N	R^a	r^b	r^c	p
AAT−	93	.49	−.48	.30	.001
AAT−	92	.42	−.02	.42	*ns*
AAT−	98	.35	−.34	.16	*ns*
TAS	40	.40	−.35	.25	.05
AAT+	93	.36	.25	.30	.05
AAT+	92	.47	.23	.42	.05
AAT+	98	.32	.28	.16	*ns*

[a] Final examination grade—Anxiety & Aptitude.
[b] Final examination grade—Anxiety.
[c] Final examination grade—Aptitude.

The second explanation holds that the significant correlation of specific anxiety scales with aptitude is really the result of a mediating process of rationalization and justification for poor performance. In order to evaluate this argument, the entering freshman class was given a specific anxiety scale. Freshmen entering Stanford have, in the majority, a past history of similar academic success experiences in high school, and they have had no past experience with the college academic performance criteria which anxiety scales were to predict. The AAT was administered to the entire freshman class at Stanford during their orientation week in the Fall of 1955. After the end of their first year, a random sample of 96 was drawn and the AAT was correlated with their grade-point averages. Correlations of .42 with the facilitating scale and −.29 with the debilitating scale were obtained. Both correlations are significant beyond the 1% level. Comparing these correlations with those obtained for a sample of 70 upperclassmen for whom exactly comparable data were on hand, no significant differences were found.

This is a weak refutation of the second explanation, at least at the college level. But, of course, college students have had anxiety experiences regarding high school achievement, and thus, it is still possible that the freshmen could be rationalizing via the AAT for their high school grades. But the two scales were correlated with high school grade-point average .29 and −.21, respectively. These correlations are both significant at the 5% level, and they are not significantly different from .42 and −.29. Thus, it seems the data cannot provide a definite answer to the issue involved—nor, probably, can any correlational study of this nature.

Third, if a timed aptitude test arouses stress cues, which elicit anxiety responses which affect the aptitude test performance, then a correlation between an anxiety measure and an aptitude measure could be interpreted to mean that the anxiety scale measures the anxiety which is a component in the aptitude test situation and which is reflected in the aptitude test scores. If this were true, then it would follow that an *aptitude* measure would correlate more highly with an anxiety-provoking college examination than with one that aroused little anxiety. Because the final examinations require more involvement on the part of the student and count more toward his total grade, these tests should be more anxiety-provoking than the midterm examination. From this reasoning, these two examinations were compared to provide the anxiety difference against which to test the aptitude measure. In no sample was there a significantly greater correlation between aptitude and final examination than between aptitude and midterm examination. These findings constitute a weak refutation of the third explanation.

A more definitive technique for evaluation of this explanation would be to administer the aptitude test to a group of students before and after they had been subjected to a course of psychotherapeutic treatment directed at the reduction of examination anxiety. If the increase in the mean aptitude score for the group receiving psychotherapy were significantly larger than any increase in the control group which had not received the psychotherapy, then we would have strong support for the third explanation of the relationship between aptitude and anxiety. This experiment would be similar to that by Pomeroy (1950), who demonstrated increases in finger-maze performance after only 10 min. of cathartic-type therapy. Until such time as appropriate data are available, however, one must hold in abeyance any final conclusions regarding the validity of this explanation.

In summary, regarding these three explanations, the first is refuted, doubt cast on the second, and the third neither refuted nor supported. Present evidence allows one to say only that either the third or the fourth explanation (that is, that the obtained correlation is a measure of the relationship between the two constructs—anxiety and aptitude) may be correct, or that both of them may be true. Had the first and second explanations been supported, they would support a preference for the use of a general anxiety scale because it is free of the contaminating influence of aptitude. The last two explanations support a preference for a specific anxiety scale because of its greater sensitivity. Since the evidence does indicate that these explanations are more probably correct, one has further support beyond that derived from the analysis of the interrelationships of the anxiety scales for the use of the specific rather than the general anxiety scales for prediction in academic performance situations.

DIRECTION OF EFFECT

A study of the anxiety literature leads to the question of how anxiety, when aroused in an examination, will affect performance; that is, whether it will facilitate it, debilitate it, or perhaps have no effect on it at all. On the basis of their earlier theoretical considerations, Sarason, Mandler, and Craighill (1952) state the following hypothesis:

> When a stimulus situation contains elements which specifically arouse test or achievement anxiety, this increase in anxiety drive will lead to poorer performance in individuals who have test-irrelevant [incompatible or interfering] anxiety responses in their response repertory. For individuals *without such response tendencies,* [italics added] these stimulus elements will raise their general drive level and result in improved performance (p. 561).

The implication is that measurable anxiety responses, when present, are debilitating to performance, an implication which is reflected in the Mandler-Sarason TAS. All the items of the TAS are unidimensional, i.e., anxiety responses are either debilitating or not (e.g., "While taking a course examination, to what extent do you worry? Worry a lot. . . [to]. . . Worry not at all." [Item No. 35]). From the absence of negative responses for an *S*, that is, a low score on the test, they infer that when anxiety-provoking cues are present in the environment, "these stimulus elements will raise [his] general drive level and result in improved performance for that *S*." (Sarason, Mandler, & Craighill, 1952, p. 561). This results in their confounding the two alternatives of a facilitating effect of anxiety and no effect of anxiety on academic performance and leads to their failure to allow for a third alternative, the possible existence of an individual whose anxiety responses in an anxiety-provoking situation do not affect test performance either by improving or by depressing the score. Such an individual would have to be described in their formulation as one whose negative or debilitating response effect is just counteracted by the increased task-performance drive.

There is another approach which does not require this unwarranted inference and according to which an anxiety scale would measure the presence and intensity of both kinds of anxiety responses, those which facilitate performance and those which interfere with it. If an individual scored low on both of these scale components, he would be considered to be unaffected (insofar as test performance was concerned) by anxiety-provoking cues. In other words, the facilitating effects of anxiety would then be measured independently and not inferred from the absence of negative responses. This independent measure would allow for the possibility of the absence of both types of response as well as for the presence of either one.

Theoretically, Mandler and Sarason would predict, on the basis of their ideas, that two such independent measures of anxiety would be highly negatively correlated. That is, if an individual had a great deal of facilitating anxiety, he would have little debilitating anxiety and vice versa. This is the assumption underlying their scale because it allows them to use only the debilitating index and to infer facilitating anxiety from the absence of debilitation. This assumption may be unnecessary. In fact, these two constructs of debilitating and facilitating anxiety may be uncorrelated. Thus, an individual may possess a large amount of both anxieties, or of one but not the other, or of none of either. The nature of this correlation can be determined empirically following the construction of two such independent measures of anxiety, a scale measuring the facilitating effects of anxiety on achievement performance, and a separate scale measuring the debilitating effects of anxiety on achievement performance.

The Achievement Anxiety Test (AAT) was constructed to make this determination. It consists of two independent scales: a facilitating scale of nine items based on a prototype of the item—"Anxiety helps me to do better during examinations and tests"; and a debilitating scale of 10 items based on a prototype of the item—"Anxiety interferes with my performance during examinations and tests." Both scales have gone through numerous revisions based upon item analyses, correlations with various criteria, and theoretical reformulations. The test-retest reliabilities for a 10-week interval are .83 and .87, respectively. The test-retest reliability over an 8-month period is .75 for the facilitating scale and .76 for the debilitating scale. The two scales are administered in one questionnaire, the items randomly mixed (as presented below). The *S*s answer each item on a five-point scale, indicating the degree to which the item applies to them. The AAT includes neutral buffer items in addition to the items listed. The numbers in the parentheses to the right are the actual item numbers on the AAT.

Facilitating Anxiety Scale

1. I work most effectively under pressure, as when the task is very important. Always—Never. (2)
2. While I may (or may not) be nervous before taking an exam, once I start, I seem to forget to be nervous. I always forget—I am always nervous during an exam. (9)
3. Nervousness while taking a test helps me do better. It never helps—It often helps. (11)
4. When I start a test, nothing is able to distract me. This is always true of me—This is not true of me. (12)
5. In courses in which the total grade is based mainly on *one* exam, I seem to do better than other people. Never—Almost always. (14)
6. I look forward to exams. Never—Always. (16)
7. Although "cramming" under pre-examination tension is not effective for most people, I find that if the need arises, I can learn material immediately before

an exam, even under considerable pressure, and successfully retain it to use on the exam. I am always able to use the "crammed" material successfully—I am never able to use the "crammed" material successfully. (19)

8. I enjoy taking a difficult exam more than an easy one. Always—Never. (21)

9. The more important the exam or test, the better I seem to do. This is true of me—This is not true of me. (24)

Debilitating Anxiety Scale

1. Nervousness while taking an exam or test hinders me from doing well. Always—Never. (1)

2. In a course where I have been doing poorly, my fear of a bad grade cuts down my efficiency. Never—Always. (3)

3. When I am poorly prepared for an exam or test, I get upset, and do less well than even my restricted knowledge should allow. This never happens to me—This practically always happens to me. (5)

4. The more important the examination, the less well I seem to do. Always—Never. (6)

5. During exams or tests, I block on questions to which I know the answers, even though I might remember them as soon as the exam is over. This always happens to me—I never block on questions to which I know the answers. (10)

6. I find that my mind goes blank at the beginning of an exam, and it takes me a few minutes before I can function. I almost always blank out at first—I never blank out at first. (15)

7. I am so tired from worrying about an exam, that I find I almost don't care how well I do by the time I start the test. I never feel this way—I almost always feel this way. (17)

8. Time pressure on an exam causes me to do worse than the rest of the group under similar conditions. Time pressure always seems to make me do worse on an exam than others—Time pressure never seems to make me do worse on an exam than others. (18)

9. I find myself reading exam questions without understanding them, and I must go back over them so that they will make sense. Never—Almost always. (23)

10. When I don't do well on a difficult item at the beginning of an exam, it tends to upset me so that I block on even easy questions later on. This never happens to me—This almost always happens to me. (26)

Part of the refinement of the scales was accomplished through means of empirical techniques. After a large number of items had been constructed for each scale, the scales were used to predict several performance criteria, such as grade-point averages and final examination grades. These data were item-analyzed to give the correlation of each item with the criterion. Those items were retained which were highly correlated with the criteria but which were not correlated with each other. In this way, it was hoped to minimize the intercorrelation of the scales without affecting their validity coefficients. The final correlations between the facilitating and debilitating scales were −.37, −.34, −.43, and −.48, drawn from four different samples. For these four samples, the average means and standard deviations for the facilitating and debilitating scales of the AAT were as follows: for the AAT+, $M=27.28$, $SD=4.27$; for the AAT−, $M=26.33$, $SD=5.33$. The average correlation for the combined N (379) was −.37. All these correlations are significant beyond the 1% level. Hence, in spite of efforts to separate the two scales empirically, a low but significant correlation remains. The point seems made, however, that the correlation is neither perfect nor high as Mandler and Sarason imply and require as an assumption for the use of their scale.

In general, the two scales of the AAT correlated about equally, though opposite in sign, with the other variables discussed in this paper (see Tables 1, 2, and 3). The important question is whether, when predicting an academic performance score such as college grade-point average, there is any advantage in using both scales rather than using merely a conventional debilitating anxiety scale such as the AAT− alone or the Mandler and Sarason TAS. These two measures of debilitating anxiety (the AAT− and the TAS) correlate .64 with an N of 40, indicating considerable equivalence. Table 6 answers the question of whether a scale which measures the facilitating effect of anxiety justifies itself. In the last two columns to Table 6 it can be seen that for three separate samples, the multiple correlations, using both the plus and minus scales to predict grade-point average, are signifi-

TABLE 6

MULTIPLE CORRELATIONS PREDICTING GRADE-POINT AVERAGE (GPA) FROM THE AAT

N	R^a	r^b	r^c	r^d	p^e	p^f
93	.50	.36	−.45	−.37	.001	.05
92	.32	.32	−.08	−.34	*ns*	.01
96	.54	.50	−.40	−.43	.05	.001

[a] GPA · AAT+ & AAT−.
[b] GPA · AAT+.
[c] GPA · AAT−.
[d] AAT− · AAT+.
[e] Significance of adding AAT−
[f] Significance of adding AAT+.

cantly better predictors than the minus or the plus scales alone.

It seems clear, therefore, that the incorporation of items designed to measure facilitating anxiety into a scale which already effectively measures debilitating anxiety can significantly increase the prediction of academic performance scores.

SUMMARY

The results support the following conclusions:

1. Specific anxiety scales (scales having items specific to the academic test situation) and general anxiety scales are measuring, to a significant extent, something different from one another.

2. The specific anxiety scales are better predictors of academic performance than are the general anxiety scales.

3. The general anxiety scales are not significantly related to verbal aptitude, while the specific anxiety scales are all related to aptitude.

4. The specific anxiety scales, although more highly correlated with aptitude than the general anxiety scales, are, nevertheless, more often than the general anxiety scales, able to account for variance in academic performance other than that accounted for by a measure of aptitude.

5. An explanation of the relationship between anxiety scales and measures of aptitude, based on the use of an anxiety scale as a vehicle for rationalization or justification by students with poor past performance in academic situations, is not supported by these data.

6. On the basis of these data, no conclusion may be reached about explanations of the relationship between anxiety scales and measures of aptitude on the assumption that aptitude scores are, themselves, affected by anxiety. A possible experiment which may clarify this issue is suggested.

A scale for the measurement of "facilitating" anxiety was constructed, which added significantly to the prediction of grade-point average when it was combined with a measure of "debilitating" anxiety. The theoretical and practical significance of these findings was discussed.

REFERENCES

ALPERT, R. Anxiety in academic achievement situations: Its measurement and relation to aptitude. Unpublished doctoral dissertation, Stanford Univer., 1957.

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

COLLEGE ENTRANCE EXAMINATION BOARD. *A description of the College Board Scholastic Aptitude Test.* Princeton: Educ. Testing Serv., 1956.

DAVIDS, A. Relations among several objective measures of anxiety under different conditions of motivation. *J. consult. Psychol.*, 1955, **19**, 275–279.

DAVIDS, A., & ERIKSEN, C. W. The relation of manifest anxiety to association productivity and intellectual attainment. *J. consult. Psychol.*, 1955, **19**, 219–222.

FREEMAN, M. J. The development of a test for the measurement of anxiety: A study of its reliability and validity. *Psychol. Monogr.*, 1953, **67**(3, Whole No. 353).

MANDLER, G. Anxiety and performance: Empirical and theoretical aspects of anxiety scales. Unpublished manuscript, Lab. Soc. Relat., Harvard Univer., 1954.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

PECK, RUTH. The influence of anxiety factors upon the effectiveness of an experimental "counselling session." Unpublished doctoral dissertation, State Univer. Iowa, 1951.

POMEROY, D. S. Ameliorative effects of "counselling" upon maze performance following experimentally induced stress. *Amer. Psychologist*, 1950, **5**, 327. (Abstract)

SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.

SCHULZ, R. E., & CALVIN, A. D. A failure to replicate the finding of a negative correlation between manifest anxiety and ACE scores. *J. consult. Psychol.*, 1955, **19**, 223–224.

SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the USC as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65–72.

(Received May 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 216–222

EFFECTS OF WAITING WITH OTHERS ON CHANGES IN LEVEL OF FELT ANXIETY[1]

LAWRENCE S. WRIGHTSMAN, JR.[2]
University of Minnesota

An old familiar saying has it that "misery loves company." This having been verified empirically by Schachter (1959), a next question might simply be "Why?" Why is it that most people, ordinarily content to be left alone, become seekers of others when they become upset, moody, or threatened? Why does an increase in a person's level of anxiety cause him to be more likely to want to be with others?

Several possible reasons for the relationship between anxiety and the affiliative tendency come easily to mind, including the expectation of irrelevant and diverting conversation with others and the idea of "safety in numbers." But a series of analytic experiments by Schachter rule out some of these as valid explanations. Schachter concludes that an anxious subject's desire to be with others is determined by (*a*) needs for direct reduction of anxiety, and/or (*b*) needs for evaluation of one's emotional state. The present study examines some of the implications of the relationship between anxiety and affiliation.

The first explanation implies that people want to be with others when anxious because they expect from others sympathy and reassurance that is directly anxiety-reducing. ("Anxiety" refers here to a feeling of uneasiness or concern about participating in an impending unpleasant event; unlike generalized anxiety or manifest anxiety it is conceived as linked to a specific situation.) According to the first hypothesis of the study, then, if persons are anxious about an impending event, being with others is more effective in reducing their anxiety than is being alone.

Despite the widely held belief in the effectiveness of the presence of other people in moderating one's extreme fears, there is no direct evidence that other people's statements facilitate the reduction of anxiety about a specific situation. Indirect evidence, however, comes from Schachter's study (1959). His subjects were given a description of the "electric shock experiment" in which they were to participate. By varying description of the anxiety-producing features of the "experiment" and the alternatives involved in a short waiting period before anticipated participation in it, Schachter found that *S*s who expected a very painful experiment more frequently chose to wait with others than did their counterparts who were led to believe that the experiment would be painless. Further evidence is afforded by the study of Davitz & Mason (1955), who found that the presence of one rat tended to reduce the strength of a fear response in another rat.

The second hypothesized explanation for the link between anxiety and affiliation implies that the presence of others gives a person a chance to evaluate his own level of emotion by comparing it with that of others. Perhaps the situation is so novel or ambiguous that a person is unclear about the appropriate emotional response. In such a case the responses of others may serve to indicate the proper level of concern.

This interpretation would seem to follow from an extension to the evaluation of emotions of the scheme developed by Festinger (1954) for explaining how certain processes of social comparison operate in the evaluation of an ability or an opinion. (The theory has been verified empirically by Dreyer, 1954; Gerard, 1953.) Festinger hypothesizes that there is a drive to evaluate one's own abilities and opinions, and when no objective method of evaluation is possible, the person uses social reality—the statements and behaviors of others—to aid in his self evaluation. For social comparison to be most informative, the person's comparison objects should be other persons whose levels of ability or opinion are similar to his own. The process of social comparison thus gives rise to pressures toward uniformity, by which the levels of others are brought closer to the person's own level.

---

[1] This study is a condensation of a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Minnesota. The writer wishes to express his appreciation to Stanley Schachter for his guidance in the study. This experiment is part of a program of research on Social Determinants of Emotional State which is being carried out at the Laboratory for Research in Social Relations under the direction of Schachter and the sponsorship of the National Institute for Mental Health of the United States Public Health Service.

[2] Now at George Peabody College for Teachers.

According to Festinger, these pressures toward uniformity result in three maneuvers: (*a*) change in the person's own ability or opinion so as to bring it closer to that of the comparison persons, (*b*) the attempt to move others toward one's own level, and (*c*) cessation of comparison of one's own position with those of persons whose abilities or opinions differ widely from one's own.

One of the purposes of this study is to determine whether the emotion of anxiety is similar to abilities and opinions in being evaluated in ways subject to influence by the responses of others. If the same social comparison processes obtain, the same maneuvers should be found in a group of people placed in a sudden anxiety-producing situation that emerge in people seeking to evaluate their opinions. The second hypothesis of this study is, then, that pressures toward uniformity in level of anxiety, namely influence processes and rejection tendencies, develop when groups of persons are placed in an ambiguous emotional situation.

But is it correct to assume that felt emotions are, in part, cognitively determined, and that there is a conceptual similarity between the determinants of emotions and the determinants of opinions? The tenor of much thinking about the emotions seems to suggest that both physiological and cognitive processes determine experienced emotions. Ruckmick (1936), for example, states that "there is a tendency . . . to admit that purely bodily expression without the cognitive or 'intellectual' factor does not qualify as an equation for the experienced real emotion," (p. 343) and that "What the child . . . feels depends more on what he cognizes or, in short, on the accumulations of his perceptual and ideational experiences," (p. 448). Young (1943) has a similar point of view. To the extent that cognitive factors do determine emotional states, emotions, like opinions, may be expected to be susceptible to social influence.

METHOD

Subjects

The subjects (*S*s) were 204 students, 120 females and 84 males, enrolled in undergraduate psychology classes. By participating the students earned bonus points applied to their examination scores.

Experimental Session

Four *S*s of the same sex were used in each experimental session. When each *S* arrived, *E* or his nurse assistant[3] placed him in a separate room in which *S* found a tray containing hypodermic needles, syringes, cotton, alcohol, and other injection equipment. After completing a few biographical questions, including several dealing with his physical health, each *S* heard over the intercommunication system a tape recording describing the experiment in which he was to participate. The experiment was described as one dealing with glucose level and mental activity. *S* was told that his glucose level was to be drastically changed by a series of injections of either a depressant drug or a glucose additive. Because these substances had to be injected directly into the bloodstream, it was likely that they would be more painful than routine injections. *S* was then warned that the injection of the depressant drug might make him feel drowsy and lacking in energy but that he should be able to respond. If he was given the additive, it would raise his glucose level to an absolute maximum amount near the point which brings on glucose-shock, fainting, and similar symptoms of hyperglycemia or diabetes. A quick blood test to determine glucose level would be given by pricking a finger. All the above operations were to occur in private.

Immediately after hearing this, each *S* was asked to indicate on a formsheet how at-ease or ill-at-ease he felt about participating in the described experiment by writing a number from 0 (completely at-ease) to 100 (completely ill-at-ease).

At this point the procedure was varied for the three conditions of the experiment. Two Together conditions were used, along with an Alone condition. Verbal communication was forbidden in one Together condition, but allowed to run its course in the other, to permit measuring the possible effects of verbal communication on reduction of anxiety. All four *S*s in a particular experimental session were in the same condition, the particular condition being determined at random.

If the *S*s were in the Together-Talk (T-T) condition, they were brought to a large waiting room and told:

> You are supposed to wait here until we are ready

---

[3] The author wishes to thank Bernice P. Flynn, who served authentically as the nurse assistant.

TABLE 1

Mean Responses and Changes in Measures of Anxiety

(Data for 17 groups in each condition, 68 *S*s in all)

Condition	Prewaiting-Period Measure	Post-waiting-Period Measure	Change from Pre- to Postwaiting-Period Measure
Together-Talk	43.7	37.3	−6.4
Together-No-Talk	42.8	36.6	−6.2
Alone	41.3	37.2	−4.1

> to use you. You can read or smoke or study if you like. If you want to talk, that's okay. You can talk about the experiment or anything you like. There will be only a short wait.

If the *S*s were to be in the Together-No-Talk (T-N-T) condition, they were brought to the same room and told:

> You are supposed to wait here until we are ready to use you. You can read or smoke or study if you like. However, please do *not* talk. Part of the experiment is of a verbal nature and we want only your own ideas to be given. Therefore, it is important that you not talk about the experiment or anything else. There will be only a short wait.

If the four *S*s were in the Alone condition, each was left in his separate room and told:

> You are supposed to wait here until we are ready to use you. You may read or smoke or study if you like. There will be only a short wait.

Immediately after this announcement, *E* and his assistant left the room for 5 min. *S*s in the Together conditions were observed through a one-way screen. All *S*s in the T-N-T condition followed the restriction against talking. After 5 min. the assistant returned and told the *S*(s), "We are about ready for you but before we begin I need to get some more information from you." She then gave each *S* a set of questions which consisted of a repetition of the prewaiting-period measure and other questions dealing with *S*'s concern about the experiment.

When they completed these, *S*s were brought together if they had been separated and a "catharsis" session was held, in which the true purpose of the experiment and the need for deception were explained. *S*s' questions were answered, and they agreed to keep the true nature of the experiment secret. Each *S* then completed a questionnaire dealing with his experimental background, his suspicions about the described experiment, and the extent of his friendship with the other *S*s at that session.

RESULTS

Effect of Being with Others on Reduction of Anxiety

If the first hypothesis is correct, *S*s waiting together should show greater reduction of anxiety than do *S*s waiting alone. Table 1 presents the mean responses for the pre- and postwaiting-period measures for *S*s in the three conditions. Means for the three conditions are quite similar on all measures. None of the differences between conditions is statistically significant, nor are differences between Alone condition *S*s and all Together condition *S*s pooled together. In each condition the average change was in a downward direction, from four to six points. When all conditions are combined, the mean level of anxiety of 37.0 on the postwaiting-period measure was significantly less ($p < .001$) than the mean of 42.5 on the same measure before the waiting period.

The number of *S*s wishing to withdraw from the experiment without receiving experimental credit was also similar in each condition. That 26 of the 203 *S*s responding to this question wished to withdraw, gives some indication of the effectiveness of the described experiment in creating anxiety.

The nonsignificant differences pertaining to the first hypothesis give no support for the belief that being with others directly reduces anxiety. But Schachter (1959) found that the relationship between anxiety and the desire to be with others was almost entirely restricted to those *S*s who were first or only children. Accordingly, such *S*s might be affected more strongly than others by the social situation. Table 2 presents the results of an analysis by birth order.

First-born *S*s in each of the Together conditions showed a greater net change than did first-born *S*s in the Alone condition. In both the Together-vs.-Alone comparisons, differences are significant at the .05 level, using a two-tailed Mann-Whitney *U* test because of the heterogeneity of variances. Thus, it appears that for *Ss who are first children or only children*, being with others is indeed effective in reducing anxiety over an impending unpleasant event.

In the case of the *S*s in Table 2 who were later children, however, the mean net change does not differ from one condition to another. No mean differences are statistically significant. For these *S*s being with others appears not to have reduced anxiety.

Table 2 introduces a new measure, that of

TABLE 2

MEAN RESPONSES ON MEASURES OF ANXIETY ACCORDING TO BIRTH ORDER OF *S*

Condition (First-born and only-child *S*s)	*N*	Prewaiting-Period Measure	Postwaiting-Period Measure	Net change, Pre- to Postwaiting Period	Gross Change, Pre- to Postwaiting Period
T-T[a]	32	39.8	31.8	−8.0[b]	10.8[c]
T-N-T[d]	33	43.2	33.9	−9.3[b]	9.9[b]
Alone[a]	41	42.1	38.6	−3.5	4.2
(Later-children *S*s)					
T-T[a]	35	47.8	42.9	−4.9	6.6
T-N-T[d]	35	42.4	39.1	−3.3	4.2
Alone[a]	26	39.9	35.4	−4.5	6.0

[a] Birth order of one *S* in T-T (Together-Talk) condition and one *S* in Alone condition were not recorded.
[b] Significantly different from comparable Alone condition mean at .05 level, using Mann-Whitney *U* test.
[c] Significantly different from comparable Alone condition mean at .01 level, using Mann-Whitney *U* test.
[d] Together-No-Talk.

"gross change," or the absolute magnitude of change disregarding direction. For this measure the same tendency is found. First-born *S*s waiting together show significantly greater gross change than first-born *S*s waiting alone; again in the case of later-children *S*s there are no differences. There is a notable difference between the first children in the T-T and T-N-T conditions: those in the T-T condition show greater gross change but less net change than do first-children *S*s in the T-N-T condition. A comparison of first borns and later children *within* the T-T condition indicates that they differ significantly in gross change but not in net change. On the other hand, in the T-N-T condition first-born *S*s show significantly greater change, both gross and net, than do later-children *S*s. It thus appears that verbal communication may actually lead to an increase in anxiety.

On the remaining postwaiting-period measures of anxiety, the between-condition differences in birth order were similar to those already presented.

Effect of Being with Others on Self-Evaluation

If others are used as objects of social comparison in a group situation, changes in a person's self-ascribed level of anxiety and influence on that of others should occur. According to the second hypothesis of this study, then, interindividual variability in level of anxiety should be reduced during the waiting period for groups of four people waiting *together*. Such homogenization should not occur in sets of four *S*s each waiting alone. The four *S*s at each session were treated as a unit, and the ratio of the range after the waiting period divided by the range before the waiting period was used as a measure of homogenization. The second hypothesis then holds that lower average range ratios should occur in the Together conditions than in the Alone condition.

The results confirm the expectation. The mean homogenization ratios in both of the Together conditions are significantly less ($p < .05$) than the mean ratio for the Alone condition (T-T, 17 sets, .83; T-N-T, .84; Alone, .98). The Alone condition shows a level of homogenization that does not differ significantly from the expected ratio of 1.0, while in the Together conditions the ratios are significantly smaller than 1.0 at the .02 level or better.

Since the mean initial ranges for the three conditions were quite similar, the differences between conditions cannot be attributed to regression effects. But in the two Together conditions other possible artifactual causes of homogenization should be considered, particularly those related to the general tendency toward reduction of anxiety and to "floor" effects associated with the zero point of the scale. This kind of artifact should have the

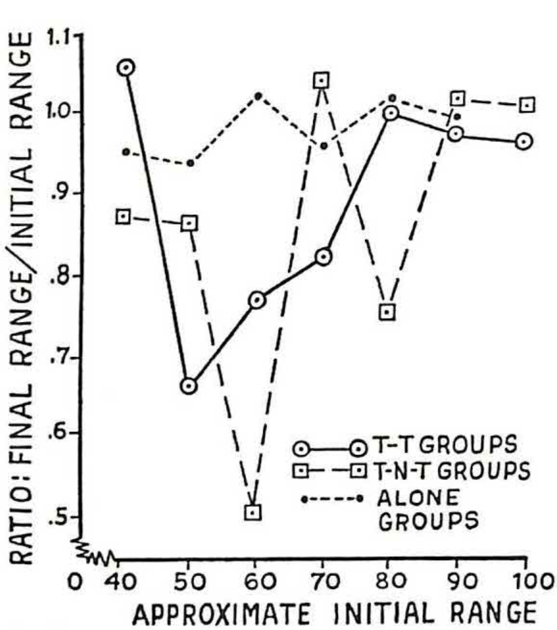

FIG. 1. The effect of initial range on extent of homogenization. (T-T refers to the Together-Talk condition; T-N-T to Together-No-Talk.)

effect that groups in which anxiety was substantially reduced should also show a high degree of homogenization; there should be a sizable correlation between the amount of reduction in anxiety and the measure of homogenization. In the T-T condition this correlation is +.31 and nonsignificant, indicating that any such artifact has a minimal effect. But in the T-N-T condition it is + .57, which is significant at the .02 level. Part of the homogenization of levels of anxiety in the T-N-T condition thus appears to be attributable to the general reduction in anxiety and the floor set by the zero point of the scale.

To the extent that interindividual variability in anxiety is nonartifactually reduced, indirect evidence is provided for the first two of Festinger's hypothesized maneuvers to reduce discrepancies—change in the person's own responses and the attempt to change the responses of others. If *S*s employ processes of social comparison to facilitate self-evaluation of their emotional states, evidence should be present in the T-T condition for Festinger's third hypothesized maneuver—cessation of comparison with those who differ widely from the person's own position. In the T-T condition, this effect should emerge more often in groups in which the initial range in levels of anxiety is great. In such groups, from the perspective of a given member, other group members should often appear so deviant and so resistant to change that they are not within the limits necessary for social comparison. There should thus be less homogenization of levels of anxiety in such groups than in the groups with moderate initial range.

Few pressures toward uniformity should be expected in T-T groups with a narrow initial range because the members are already quite uniform in level of anxiety; less homogenization of levels of anxiety may therefore be expected in such groups than in those with a moderate initial range. In groups under the Alone condition, on the other hand, homogenization should be unrelated to initial range since there is no opportunity for social comparison to occur. The T-N-T condition is more difficult to predict but should resemble the T-T-condition results to some degree if processes of social comparison are operating.

Figure 1 indicates that the data are in agreement with these predictions. (Because the initial range for most groups was a multiple of five, ranges have been grouped together in Fig. 1. From one to five groups are represented by each point.) The three T-T groups with initial ranges of 90 points or greater and the two groups with an initial range of 40 points or less showed less homogenization than did the 12 groups with moderate initial ranges. (Of the latter groups, five had a range of approximately 50 points, one 60 points, five about 70 points, and one of 80 points.) No such pattern is evident for the sets in the Alone condition, where homogenization is unrelated to initial range. The results for the T-N-T condition were more erratic, but their resemblance to those for the T-T condition tentatively suggests that comparison processes may be active in the absence of verbal communication.

A possible artifactual cause for the differences in homogenization between wide-range and moderate-range groups might lie in a tendency for those *S*s whose initial reaction was extreme (either less than 25 or greater than 75) to be more resistive to change, but a further analysis of the data gives no support to this possibility.

DISCUSSION

Effect of Being with Others on Reduction of Anxiety

The differences according to birth order in the anxiety-reducing effect of being with others fit in neatly with Schachter's finding (1959) that first-born *S*s have a greater desire to be with others when anxious. It now appears that their reason for this greater desire can be attributed to the fact that being with others directly reduces their anxiety. Later-children *S*s, for whom being with others does not have this effect, therefore have no particular desire to be with others when anxious.

The results for the T-N-T condition are consistent with another of Schachter's findings, that first-born *S*s desire to be with others when anxious even when verbal communication is not permitted. The present study finds, correspondingly, that when such *S*s *are* placed in a situation without verbal communication, where only the sheer physical presence of others is available, they respond with a greater reduction in anxiety than occurs in similar *S*s who remain alone.

But why this birth-order effect? The cause is certainly not a contemporary one; it must be traced to childhood experiences. There are indications in the literature that first and later children differ in various ways in their early environment. Studies of the parental treatment of different siblings (Gesell & Ilg, 1942; Sears, Maccoby, & Levin, 1957) indicate that the first child receives more attention and nurturance from his parents than do his later siblings. Sibling rivalry, with the displacement of hostilty upon the younger child (Baldwin, 1955), appears to occur frequently. On the basis of present knowledge, however, there is no compelling reason for preferring any one of several possible explanations, which invite further research.

Effect of Being with Others on Self-Evaluation

The strong support given to the second hypothesis indicates that Festinger's scheme for explaining self-evaluation through a social comparison process is useful for the emotion of anxiety as well as for opinions and abilities. Can it then be concluded that because cognitive factors have some effect in determining this particular emotional state that the emotions in general are also susceptible to social influence? Some evidence for an affirmative answer is found in a study which demonstrates that another emotion—the state of boredom or satiation—can be influenced by processes of social comparison (Horwitz, Exline, Goldman, & Lee, 1953). In this investigation of the effects of different group characteristics on task satiation, it was found that in the one condition where verbal communication was present, the variability among the members' levels of satiation was much less than the variability within the no-communication conditions.

SUMMARY

Two determinants for the relationship between anxiety and the affiliative tendency have been proposed. These are the need for direct reduction of anxiety and the need for self-evaluation of level of anxiety. The purpose of this study was to examine some of the implications of these two explanations. It was hypothesized that (*a*) being with others would be more effective in reducing anxiety than would being alone, and (*b*) pressures toward uniformity in level of anxiety would be elicited in groups of persons placed in an ambiguous emotional situation, as a consequence of requirements for self-evaluation of level of anxiety.

The first hypothesis was not confirmed, but an alternative hypothesis—that for first-born *S*s, being with others is more effective in reducing anxiety than is being alone—received strong support.

The second hypothesis received strong support. Pressures toward uniformity in regard to level of anxiety emerged in *S*s waiting together. It was concluded that a person does evaluate his level of anxiety through the process of social comparison and that level of anxiety shares with opinion the feature of being partially determined by cognitive processes.

REFERENCES

BALDWIN, A. L. *Behavior and development in childhood.* New York: Dryden, 1955.

DAVITZ, J. R., & MASON, D. J. Socially facilitated reduction of a fear response in rats. *J. comp. physiol. Psychol.*, 1955, **48**, 149–151.

DREYER, A. S. Aspiration behavior as influenced by expectation and group comparison. *Hum. Relat.*, 1954, **7**, 175–190.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.

GERARD, H. B. The effects of different dimensions of disagreement in the communication process in small groups. *Hum. Relat.*, 1953, **6**, 249–272.

GESELL, A., & ILG, FRANCES L. *Infant and child in the culture today: The guidance and development in home and nursery school.* New York: Harper, 1942.

HORWITZ, M., EXLINE, R. V., GOLDMAN, M., & LEE, F. J. Motivational effects of alternative decision-making processes in groups. *ONR tech. Rep.*, 1953. (Contract N6ori-07144, Group Psychol. Branch)

RUCKMICK, C. A. *The psychology of feeling and emotion.* New York: McGraw-Hill, 1936.

SCHACHTER, S. *The psychology of affiliation.* Stanford, Calif.: Stanford Univer. Press, 1959.

SEARS, R. R., MACCOBY, ELEANOR, & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

YOUNG, P. T. *Emotion in man and animal.* New York: Wiley, 1943.

(Received June 24, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 223-230

THE INFLUENCE OF SOME TYPES OF POWER RELATIONSHIPS AND GAME STRATEGIES UPON THE DEVELOPMENT OF INTERPERSONAL TRUST[1]

LEONARD SOLOMON[2]

Research Center for Human Relations, New York University

THIS study is part of a research program developed by Deutsch (1957), which sought to investigate the personal and situational determinants of interpersonal trust. In his initial studies Deutsch examined the influence of cooperative, individualistic, and competitive motivational orientations upon the development of mutual trust in a two-person bargaining game. A bargaining relationship is one in which the participants have a conflict of interests yet are dependent upon each other in order to avoid a mutually detrimental outcome (Schelling, 1958). One rather striking finding to emerge from Deutsch's study was the fact that under conditions of noncommunication the individualistically oriented dyads did not differ from the competitive ones with respect to the number of trusting choices made in a ten-trial game. Contrary to what would have been predicted on a common sense basis, the individualistic subjects (*S*s) were unable to act rationally in the absence of mutual trust. Each player was unable to subordinate immediate self-interest for long range group gain and the relationship tended to deteriorate into a self-defeating competition. These findings stimulated several follow-up studies that attempted to determine the minimal social or situational conditions under which mutual trust could arise in dyads with an individualistic orientation. Thus, Loomis (1959) examined the effects of variations in the opportunity to communicate while Farr (1957) dealt with the influence of a mutually disliked third person upon the development of mutual trust.

The present study sought to determine how a player in different positions of relative power could exercise this power in the form of various game playing strategies so as to induce an individualistic *S* to adopt a cooperative orientation. Deutsch's formulations concerning the situational conditions necessary for the realization of mutual trust, along with Heider's principle of cognitive balance (Heider, 1958) provided a theoretical basis for the derivation of several hypotheses.

Trust and suspicion. For the purposes of economy and clarity of exposition, the symbol *S* is used to refer to the perceiver (also the subject) and his experiences or cognitions; *O* refers to another person whose action affects *S*. In the terms of the dyadic game employed in this study, *S* refers to Player 1 and *O* refers to Player 2.

Deutsch (1958) has proposed the following conceptualizations of interpersonal trust and suspicion:

> A person (*S*) may be said to have trust in the occurrence of an event, if *S* expects its occurrence and his expectation leads to behavior that he perceives to have positive motivational consequences if the expectation is confirmed and negative motivational consequences if it is not. *S* may be said to be suspicious of the occurrence of an event if the disconfirmation of the expectation of the event's occurrence is preferred to its confirmation, and if the expectation of its occurrence leads to behavior that is intended to reduce its negative motivational consequences.
>
> Interpersonal trust or suspicion exists when *S* trusts or suspects that *O* will do something and he perceives that *O* is aware that his behavior has relevance for *S*. Trust refers to expectations of benevolence; whereas

---

[1] This paper is adapted from a doctoral dissertation completed in 1957 at New York University. The author is deeply indebted to Morton Deutsch, under whose guidance this study was performed. The author also wishes to thank Robert S. Lee of New York University and Herbert E. Spohn, Harold Wilensky, and Jacob Cohen of the Montrose VAH research staff for their valuable suggestions with respect to the manuscript. This experiment was carried out as part of a research program on Conditions Affecting Cooperation and was supported by a grant from the Office of Naval Research under Contract NONR-285 (10). The United States government is authorized to reprint this article in part or whole.

[2] Now at Boston University.

suspicion refers to expectations of malevolence. Deutsch (1958) exemplifies these notions when he states (paraphrased):

If a mother trusts a baby sitter enough to leave her baby in her care and the sitter does not live up to this trust, the mother suffers an unpleasant consequence—harm to her baby. The trusting mother, when her trust is fulfilled, is in a more advantageous position than if she had not trusted. However, the disadvantages of trusting and having her trust unfulfilled are considerably greater than the advantages of being able to do things she could otherwise not do.

Cognitive balance. Heider (1958) has proposed that the process of social perception is steered by a need to achieve cognitively balanced states. The need for cognitive balance has been defined as a tendency for the person to cognize various aspects of his environment in such a way that the behavioral implications of his cognitions are not in contradiction. If an imbalanced state exists, the person will attempt to reduce or remove the imbalance through either cognitive reorganization or behavior. This principle and several of its derivations have been elaborated and tested by Festinger (1957) and Deutsch and Solomon (1959).

Interpersonal power. Heider has defined social power as the ability to produce intentionally events that have a particular degree of positive or negative motivational consequences for a recipient. In line with this conception the interpersonal power balance has been defined in this paper in the following ratio form: *O*'s ability to produce motivational consequences for *S* relative to *S*'s ability to produce motivational consequences for *O*. The nature of the events to be produced will determine what dimensions of ability required (i.e., physical strength or intelligence).

Game strategies. The two person bargaining games are so contrived that they represent a mixture of mutual and divergent interests. In such relationships the term *game strategy* is used to refer to the pattern of *O*'s choice behavior as it affects the preferences of *O* and *S*. As a first approximation, situations which result in both lesser loss as well as greater gain are treated, hereafter, as involving gain; loss is extended to include lesser gain as well as greater loss. The two modal forms of game strategy dealt with in this experiment are the Cooperative Strategy wherein *O*'s choice pattern promotes the maximum mutual gain, and the Noncooperative Strategy wherein *O* blocks the attainment of maximal mutual gain. The motivational source of such behavior patterns may vary. For example, *O*'s cooperation could stem from fear of retaliation (individualism) or a sincere liking for *O* (altruism). A guiding conception of this study is that the nature of the power relationship influences the kind of motive *S* attributes to account for *O*'s game strategy.

In terms of the game, *O*'s power over *S* is a reflection of his ability to achieve a preferred outcome relatively independent of *S*'s cooperation. Thus, if *O* had absolute power, he could employ a Noncooperative Strategy to obtain his optimal gain whereas if the players had equal power, *O* would have to reckon with potential losses derivable from *S*'s use of his retaliatory power. This reasoning suggests a more general proposition; the greater *S*'s power relative to *O* the more likely is *S* to expect cooperative as opposed to noncooperative behavior and to behave with trust rather than suspicion. In addition, the principle of cognitive balance serves to predict that an *S* who receives a Cooperative Strategy should tend to develop a liking for *O*, whereas, the contrary would be true if *S* received a Noncooperative Strategy. The intensity of the like or dislike is expected to be influenced by the particular power relationship that exists between *S* and *O*, or to put it another way, the perceived cost of cooperation or noncooperation to *O* influences *S*'s motive attribution and subsequent feelings toward *O*.

The specific hypotheses associated with this rationale are elaborated in the presentation of the results.

METHOD

A laboratory experiment was designed employing a two-person nonzero-sum game (see Fig. 1) in which the gains or losses incurred by each player were a joint function of the choices made between two alternatives by both players. Two sets of independent variables, namely, four types of power relationships and three types of game strategies, were created experimentally and analyzed as a 4 × 3 factorial design. A total of 256 undergraduate Queens College students served as *S*s. Thirty-six experimental groups each composed of

O in Absolute Power 1 (AP_1)
(Coordinated Preferences)

	A	B
X	(+30, +30)	(−30, +30)
Y	(+20, +30)	(−10, +30)

O in Partial Power (PP)

	A	B
X	(+30, +30)	(−30, +40)
Y	(+20, +10)	(−10, +20)

O in Absolute Power 2 (AP_2)
(Divergent Preferences)

	A	B
X	(+30, +30)	(−30, +40)
Y	(+20, +30)	(−10, +40)

O in Equal Power (EP)

	A	B
X	(+30, +30)	(−30, +40)
Y	(+40, −30)	(−20, −20)

(*S*) chooses between Rows X and Y, whereas (*O*) chooses between columns A and B. *S*'s payoffs are the first numbers in the parentheses; *O*'s are the second numbers.

The EP game matrix is similar in structure to that used by Deutsch (1957) in the study previously described.

FIG. 1. The game matrices for four power conditions.

from four to eight *S*s were randomly assigned to one of the four power conditions. Each *S* within a particular group was then exposed (by a random procedure) to one of three game strategies from his partner.

Procedure

Laboratory setting. Once in the laboratory, the *S*s were placed in cubicles so that they were unable to determine the identity of their game partners. In order to communicate their game choices, the *S*s used an electric signal switch. The two switches on each *S*'s box were wired separately to the *E*'s master control panel, which was hidden from the *S*s. The wiring scheme permitted the *E* to receive information separately from each *S* as well as to send preprogrammed choices to all *S*s simultaneously. At the outset of the experiment, the *S*s were given a standard set of instructions which are outlined below:

> You are going to participate in a study concerned with game playing behavior. You will be playing with another person whom you cannot see or communicate with. One of you will be called Person 1 (*S*), the other of the pair will be Person 2 (*O*). In this game you can either win money or lose money. This money will be *imaginary*, but we want you to feel and act as if it were real money.

Creation of power conditions. Groups of *S*s were randomly assigned to one of four game matrices, each corresponding to a particular power condition (Fig. 1). The power conditions are defined from the point of view of *O*'s position:

O in Absolute Power 1 (AP 1). O (Person 2) has absolute power over *S*. *O* can determine whether *S* wins (Column A) or loses (Column B). *S* however cannot affect *O*'s payoffs; *O* wins $30 for any combination of choices. *O* and *S*'s preferences are thus *coordinated.*

O in Absolute Power 2 (AP 2). This condition is the same as AP 1 with one difference; *O* can win more by choosing B ($40) rather than A ($30). If, however, *O* does choose B, *S* must lose. *O* and *S*'s preferences are *divergent.*

O in Partial Power (PP). O can determine whether *S* wins or loses. *S* in this condition can, however, exert some control over *O*'s payoffs. By choosing Row Y rather than Row X, *S* forces *O* into a position whereby he can win only $10 or $20, rather than a possible $30 or $40. This condition differs from AP 2 only with respect to the fact that *S* has retaliatory power in the Partial Power Condition.

O in Equal Power (EP). Both players are equally powerful in terms of the extent to which each player can determine the gains or the losses of the other. Each player can achieve his maximal gain by a choice which is detrimental to the interests of the other. (Row Y for *S* and Column B for *O*.) If a YB combination does occur then both players lose. The EP condition differs from the PP condition for *S* in two important respects: (*a*) *S* can exert a greater control over *O*'s payoffs and; (*b*) *S* has greater incentive to choose Y since he can maximize his gain ($40). Under PP a choice of X would enable *S* to maximize his gain ($30).

Game playing instructions. Regardless of the particular matrix utilized, the social implications of any combination of choices were explained to the *S*s in the following manner (illustration is based upon the EP condition):

> Person 1 chooses first between Row X or Y for each trial. Person 2 knowing Person 1's choice then makes his choice of Column A or B. (Before each of you makes his choice you will answer two questions on your game score form.) If Person 1 were to choose X he could either win $30 or $40 depending upon what Person 2 decides to choose. If Person 2 chooses A, both players win $30 whereas if he chooses B, Person 2 would win $40 while Person 1 loses $30. If Person 1 chose Y he could win $40 or lose $20. Here if Person 2 chose A he would lose $30 (Person 1 wins $40) whereas, if he chose B, both would lose $20.

All groups were given the same instructions regarding their goal in playing the game. They were instructed as follows:

> This is the orientation, or the attitude you should

have in playing this game. *If you are Person 1*(S), *your motivation should be to win as much money as you can for yourself*. You are not interested in what happens to the other person. You are not out to help the other person or to beat him. You are merely out to win as much as you can and lose as little as possible and you should feel it's important for you to do so. *If you are Person 2* (O), *you can decide for yourself just how and for what purpose you want to play the game.*

A basic requirement of the experimental situation was that the *S*s understand the mechanics of the game playing procedure as well as the interpersonal effects of their game choices. A twelve-item test was administered to the *S*s just before the game to check on the adequacy of their game understanding. Any *S* who made more than one error on this test was dropped from the experimental sample. Only nine *S*s (4%) failed by this criterion.

Creation of three game strategies. The *E* fed back a preprogrammed series of choices to the *S*s in his role as Person 2. Before describing the game strategies, it is necessary to list and define the basic types of game choices and their social implications:

Cooperative Choice (*coop. choice*) refers to a choice of X or A which serves to maximize the mutual gain.

Noncooperative Choice (*noncoop. choice*) refers to a choice of Y or B which serves to maximize the recipient's loss and minimizes the combined gain of the dyad.

Trusting Behavior refers to a coop. choice based upon expectation of receiving a coop. choice from the other player (*O*).

Suspicious Behavior refers to a noncoop. choice based upon an expectation of receiving a noncoop. choice from *O*.

Exploitative Behavior refers to a noncoop. choice based upon an expectation of receiving a coop. choice from *O*. (EP condition only)

TABLE 1

PERCENTAGE OF *S*s WITH TRUSTING AND SUSPICIOUS CHOICES IN EACH POWER CONDITION

(Trial 1)

	Absolute Power 1	Absolute Power 2	Partial Power	Equal Power
Trust[a]	43.1	22.0	30.9	24.2
Suspicion[b]	27.6	39.0	32.4	45.2
Residual[c]	29.3	39.0	36.7	30.7
	100%	100%	100%	100%
	$N = 58$	$N = 59$	$N = 68$	$N = 62$
		χ^2	*df*	*p* (1 tail)
Hypothesis 1	$AP_2 > AP_1$	5.92	2	<.03
Hypothesis 2	$AP_2 > PP$	—	2	*ns*
	PP vs. EP	—	2	*ns*

[a] *S* expects a coop. choice from *O* and makes a coop. choice.

[b] *S* expects a noncoop. choice and makes a noncoop. choice.

[c] Residual category includes both of the following:

(*a*) *S* expects a coop. choice and makes a noncoop. choice (exploitative for the EP condition).

(*b*) *S* expects a noncoop. choice and makes a coop. choice (risk taking).

Risk-Taking Behavior refers to a coop. choice based upon an expectation of receiving a noncoop. choice from *O*.

Each *S* received one of three types of game strategies from *O*:

1. *An Unconditionally Cooperative Strategy: O* makes a coop. choice (A) on every trial regardless of what *S* chooses.

2. *A Conditionally Cooperative Strategy:* On the first trial *O* makes a coop. choice. However, for each trial thereafter *O*'s choice is *reciprocal* to that of *S O* makes a coop. choice only when he receives one from *S*.

3. *A Noncooperative Strategy: O* makes a noncoop. choice (B) for every trial regardless of what *S* chooses.

After the data were analyzed, it was discovered that because some *S*s consistently made a cooperative offering in the face of a prescheduled cooperative strategy, they had the experience of mutual cooperation for six trials. These *S*s, therefore, satisfied the requirements for assignment to both the Conditional and the Unconditionally Cooperative Strategy. Moreover, they represented a select group in terms of personality predispositions as well as game experience. For these reasons this group was partialled out for separate analysis and labeled the *Residual Cooperative Group*.

Reverse trial and postgame questionnaire. After six trials of the game, *S*s were informed that for the seventh trial their roles would be reversed. *S* would now choose last. *E* "fed" *S*s in all conditions a coop. choice for the Reverse Trial. At the close of the game, *S*s were asked to complete a postgame questionnaire that tapped their attitudes toward *O* and interpretations of his game playing patterns.

Informal interviews with a number of the *S*s after they played the game revealed that they became quite involved in the game interaction. Several demanded to know the identity of their partners whom they felt had "exploited" or "double-crossed" them.

RESULTS AND DISCUSSION

Hypothesis 1. The greater the gain *O* can derive from noncooperative as compared to cooperative behavior, the more likely is *S* to behave suspiciously.

Hypothesis 2. The greater the power *O* has relative to *S*, the more likely is *S* to behave suspiciously.

In order to capture the effects of differences in power relationships in their purest form, the percentage of *S*s who were trusting or suspicious at the outset of the game (Trial 1) were compared for the four power conditions. These data are presented in Table 1.

Hypothesis 1 receives support by virtue of the fact that a significantly greater proportion of *S*s act suspiciously when initiating an interaction with an Absolutely Powerful *O* who stood to gain from a noncoop. choice as compared to one who did not ($AP_2 > AP_1$). This prediction is also supported by the finding that *S*s make a significantly greater mean number of suspicious choices during the game (Trials 2–6) in the AP_2 as compared to the AP_1 condition (see Table 2).

The test of Hypothesis 2 was based upon a comparison of *S*'s behavior when confronted with an *O* who has Absolute Power (AP_2) as compared to where *O* has Partial Power and *S* could retaliate. Although Table 1 indicates that the differences in the percentage of suspicious choices for Trial 1 were in the predicted direction, they did not attain statistical significance. Table 2, however, shows that the *S*s made a significantly lower mean number of suspicious choices in a relationship where they could retaliate (PP) as compared to one where they could not (AP_2). These findings (Trials 2–6) taken together with the first trial data substantiate Hypothesis 2.

Of interest is a comparison of the Equal Power with the Partial Power Condition for Trial 1. The EP conditions differ from the PP condition in two respects: although *S* can exert greater control over *O*'s outcome this is accompanied by a greater incentive to make a noncoop. choice (see Fig. 1). The absence of a reliable difference between the EP and PP conditions suggests that an increase in *S*'s relative power does not necessarily increase the probability of trusting behavior if the heightened power is accompanied by an inducement to be noncooperative. (See Table 1.)

Hypothesis 3. Under the conditions where *S* receives a Cooperative as compared with a Noncooperative Strategy: (*a*) *S* is more likely to trust *O*; (*b*) *S* tends to develop a greater liking for *O*.

Hypotheses 3a and 3b were tested by applying an analysis of variance to test the effects of four power conditions and four game strategies upon the development of trust and sentiment. With respect to the mean number of trusting choices made during the game (Trials 2–6) the only significant *F* ratio was derived from the main effects of game strategies ($p < .01$). Postmortem *t* tests (Snedecor, 1956) support Hypothesis 3a in that any type of Cooperative Strategy employed by *O* induced a significantly greater number of trusting choices than the Noncooperative Strategy. There was no reliable difference between the Conditional and Unconditionally Cooperative Strategy.

TABLE 2

MEAN FREQUENCY OF TRUSTING AND SUSPICIOUS CHOICES FOR THE COOPERATIVE STRATEGIES UNDER THREE POWER CONDITIONS

(Trials 2–6)

	Means			*t*	
	AP_1	AP_2	PP	AP_1 vs. AP_2	AP_2 vs. PP
Trust	2.80	2.26	2.83	1.33*	1.31*
Suspicion	0.97	1.50	0.85	1.75*	2.20*

* $p < .05$, one-tailed test.

The index of sentiment was based upon the mean of *S*'s responses to nine evaluative items of the Osgood Semantic Differential (Osgood, Suci, & Tannenbaum, 1957). The scores ranged from 1 (dislike) to 7 (like). An analysis of variance of the mean sentiment score for the 16 experimental groups yielded a significant *F* ratio for the main effects of game strategies ($F = 43.7$, $p < .001$) as well as a significant interaction effect. (Power × Strategies, $F = 2.1$, $p < .01$) Postmortem *t* tests indicated, as predicted by Hypothesis 3b, that any type of Cooperative as compared to Noncooperative Strategy induced a greater liking for *O*.

An inspection of the 16 cells suggested that the interaction effect was a function of the reversals in the direction of the differences between the two types of Cooperative Strategies. Under the EP condition the Conditional Cooperative group had a higher mean sentiment score than the Unconditionally Cooperative group; however, under the Unequal Power Conditions (AP_1, AP_2, PP) this difference reversed itself. A *t* test of the mean difference between these two pairs of means yielded a significant interaction ($t = 2.21$, $df = 41$, $p < .02$). Thus, a Conditionally Cooperative *O* was evaluated more favorably than an Unconditionally Cooperative *O* if *S* had Equal Power; however, the Unconditionally Cooperative *O* was evaluated more favorably under conditions where *S* was *dependent* upon *O*.

A behavioral index of sentiment was available through the fact that on the final trial (#7) *S* and *O* reversed roles so that *O*

TABLE 3

MEAN NUMBER OF COOPERATIVE CHOICES FOR THREE GAME STRATEGIES

(Equal Power Condition)

Strategies	Means	Groups Compared	*t* test
	Cooperative Choices		*p*(2 tailed)
Cond. Coop.	3.47	Cond. vs. Uncond. Coop.	$<.01$
Uncond. Coop.	1.43	Cond. vs. Noncoop.	$<.01$
Noncoop.	1.00	Uncond. vs. Noncoop.	*ns*

TABLE 4

THE RATIO OF EXPLOITATIVE/NONCOOPERATIVE CHOICES FOR THREE MOTIVATIONAL STRATEGIES

(Equal Power Condition)

Exploitative / All-Noncooperative[a]	Noncooperative	Unconditional Cooperative	Conditional Cooperative
Ratio = 0	20	7	15
Ratio > 0	2	7	4
p Value[b]			
Uncond. vs. Cond. Coop.		$<.01$	
Uncond. vs. Noncoop.		$<.01$	
Cond. vs. Noncoop.		*ns*	

[a] The ratio of exploitative/noncoop. choices is based upon the proportion of $\frac{\text{Y choices and A expectation}}{\text{Y choice with A or B expectation}}$ for each *S*. The distributions were highly skewed, thus a dichotomy at *O* and >*O* was made.

[b] Fisher exact test.

chose first with a programmed coop. choice and *S* chose last. In this situation the *S*s (under conditions AP_2, PP, and EP) were provided with an opportunity of gaining an additional \$10 at a cost of \$30 to *O* (noncoop. choice) or of accepting a lesser gain (−\$10) thereby enabling *O* to win \$30 (coop. choice). In effect, a coop. choice at this point represents an abandonment of a purely individualistic orientation. The results, in terms of the proportion of coop. choices made under each strategy are as follows: Residual Cooperative, 88%; Conditional Cooperative, 52%; Unconditional Cooperative, 39%; and Noncooperative, 21%. In line with Hypothesis 3 each Cooperative Strategy differed significantly from the Noncooperative Strategy ($p < .05$; cells corrected for unequal n). There were no significant differences between the Conditional and the Unconditional Cooperative Strategy.

Hypothesis 4. Under conditions where *S* and *O* are in an Equal Power relationship and can achieve their respective maximal gain via noncooperative behavior, the application of a Conditional rather than an Unconditionally Cooperative Strategy is more effective in inducing *S* to play cooperatively. Hypothesis 4 presumes that *S* learns that *O*'s cooperation is not contingent upon his own cooperation when playing with an Unconditionally Cooperative *O*, whereas in playing with a Conditionally Cooperative *O* he learns that his own cooperation is necessary in order to elicit *O*'s cooperation.

The data in Table 3 corroborates Hypothesis 4. *S*s chose a significantly greater number of coop. choices when playing with a Conditional as compared to an Unconditionally Cooperative *O*. There was no significant difference, however, between the mean number of coop. choices between the Unconditionally Cooperative and the Noncooperative Strategy. By teasing out those noncoop. choices which were based upon an expectation of cooperation (exploitation) from those based upon an expectation of noncooperation (suspicion) it was possible to make a precise analysis of *S*'s intentions.

Table 4 presents the proportion of noncoop. choices based upon an exploitative intention for the three types of game strategies. The results indicated that the *S*'s noncoop. choices are more likely to be based upon an exploitative intention when playing with an Unconditionally Cooperative *O* as compared with a Noncooperative or Conditionally Cooperative *O*. Apparently *S* was quick to learn that he could satisfy his individualistic goals most efficiently by exploiting *O*'s rigid pattern of cooperation. In spite of the fact that *S*s in this condition won more money they did not show as great a liking for the Unconditional as compared to the Conditionally Cooperative *O*. How can this finding, which on the surface contradicts reinforcement theory as well as the notion of cognitive balance be explained?

S is confronted with a pair of inconsistent cognitions; *O* (who is presumed to like himself) consistently rewards *S*'s exploitative behavior. *O*, in effect, acts in a "self-defeating" manner. *S* can resolve the cognitive imbalance in one of three basic ways: by disengaging *O* from any responsibility for his self-defeating behavior; by reinterpreting the meaning of the loss for *O*; or by inverting the tacit assumption that *O* likes himself to *O* dislikes himself. The data on motive attribution serves to shed some light on the nature of *S*'s cognitive reorganization.

S's motive attribution. Data pertaining to the type of motive *S* attributes to *O* are derived from *S*'s response to Item 3 of the Postgame Questionnaire which asked: "What do you think *O*'s purpose or goal was in playing the game in this manner?" *S*s could select from one of four fixed alternatives or choose to write in their own response. (11% of the *S*s chose the open-ended alternative.) All of the *S*s' responses were classified into one of the three following categories: (*a*) a *Cooperative Motive* defined as *O* seeks to maximize the mutual gain, (*b*) an *Individualistic Motive* defined as *O* seeks to maximize his own gain and/or *O* seeks to outdo *S*, and (*c*) an *Out of the Field Motive* defined as *O* does not wish to win money or *O* does not understand the game. Table 5 presents the proportion of *S*s who attribute each of these three motives for two types of Cooperative Strategies and two types of Power Conditions. It shows that where *S* has equal power the Unconditionally Cooperative Strategy is more likely to induce the attribution of an Out of the Field Motive, whereas, under conditions where *S* is dependent upon *O*, this strategy is more likely to induce *S* to attribute a Cooperative Motive.

Thus, an *S* who interacts with an Unconditionally Cooperative *O* (EP) resolves his incompatible impressions of *O* by attributing to *O* a disinterest in the goal or sheer cognitive confusion. In effect, the meaning of the loss for *O* has been reinterpreted, and/or *O* is disconnected from responsibility for his behavior. These findings also serve to parallel and complement the results with respect to *S*'s sentiment toward *O*. On theoretical grounds we would expect *S* to have a greater liking for an *O* whose coop. choices stem from an intrinsic desire for group gain than for an *O* whose coop. choices are accidental or only a derivative of a more primary self-centered purpose.

TABLE 5

THE PERCENTAGE OF *S*s WHO ATTRIBUTE EACH OF THREE MOTIVES TO *O* UNDER TWO POWER CONDITIONS AND TWO GAME STRATEGIES

O's Motive	Equal Power Condition[a]		Unequal Power Condition[b]	
	Conditional Cooperative ($N = 20$)	Unconditional Cooperative ($N = 14$)	Conditional Cooperative ($N = 43$)	Unconditional Cooperative ($N = 32$)
% Cooperative	75.0	57.1	52.4	100.0
% Individualistic	25.0	7.1	35.7	00.0
% Out of Field	00.0	35.8	11.9	00.0
	100.0%	100.0%	100.0%	100.0%

[a] In Equal Power Condition $\chi^2 = 9.01$; $df\ 2$; $p < .02$.[c]

[b] In Unequal Power Condition $\chi^2 = .9$; $df\ 2$; $p < .01$.

[c] Although the requirements of expected cell frequencies of 5 for using χ^2 were not met, the p value obtained suggests that the null hypothesis may be rejected.

SUMMARY

A two-person nonzero-sum game was devised to test some hypotheses regarding the influence of differences in social power and the strategy with which it is used upon the development of interpersonal trust. One member of the dyad was an *S* while the other was a programmed experimental confederate (*O*). *O* interacted with *S* in one of three conditions of relative power: *O* in Absolute Power, Partial Power, and Equal Power. *O* employed one of three types of game strategies in each power condition; Conditional Cooperation, Unconditional Cooperation, and Noncooperation.

The major findings supported Deutsch's theory of trust. They were the following:

1. *S* is more likely to engage in trusting behavior the greater the power he has relative to *O*.

2. Under conditions of Equal Power, *S* tends to respond to an Unconditionally Cooperative *O* with exploitative game behavior and less liking, whereas he tends to be cooperative and have a greater liking for the Conditionally Cooperative *O*. This difference reversed itself under the Unequal Power Conditions.

These findings are of particular relevance to the problems inherent in the socialization of a child or a patient and they suggest some of the conditions under which exploitative or altruistic motives may develop. More specifically this experiment demonstrates that a Conditionally Cooperative Strategy is most likely to facilitate the transformation of an individualistic to a cooperative orientation, particularly under conditions where there is a high degree of mutual interdependence.

REFERENCES

DEUTSCH, M. Conditions affecting cooperation. *USN tech. Rep.*, 1957, Contract NONR-285 (10).

DEUTSCH, M. Trust and suspicion. *J. conflict Resolut.*, 1958, **4**, 265–279.

DEUTSCH, M., & SOLOMON, L. Reactions to evaluations of others as influenced by self-evaluation. *Sociometry*, 1959, 22, **2**, 93–111.

FARR, J. *The effects of a disliked third person upon the development of mutual trust.* Paper read at APA, New York, September 1957.

FESTINGER, L. *A theory of cognitive dissonance.* White Plains: Row, Peterson, 1957.

HEIDER, F. *Psychology of interpersonal relations.* New York: Wiley, 1958.

LOOMIS, J. Communication, the development of trust, and cooperative behavior. *Hum. Relat.*, 1959, **4**, 305–315.

OSGOOD, C., SUCI, G., & TANNENBAUM, P. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

SCHELLING, T. C. The strategy of conflict: Prospectus for a reorientation of game theory. *J. conflict Resolut.* 1958, **3**, 203–264.

SNEDECOR, G. *Statistical methods.* Ames: Iowa State Coll. Press, 1956.

(Received July 15, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 231-238

SOCIAL EFFECTIVENESS AND SYMPTOMATIC BEHAVIORS

EDWARD ZIGLER[1]
Yale University

AND LESLIE PHILLIPS
Worcester State Hospital and Clark University

IN A RECENT study, Zigler and Phillips (1959) investigated the relationship between mental disorder and case history data. Employing the variables of age, intelligence, education, occupation, employment history, and marital status as referrents of adaptive adequacy, they examined the differences between a state hospital population and the normal population of the same geographical region. They also investigated how each of the major groups of functional disorders differed from the normal population and from one another on these variables.

Despite the limitations incurred through the use of a restricted hospital population and the employment of as unreliable an index of deviant behavior as an individual's diagnosis, Zigler and Phillips found that: (*a*) the hospitalized individual tends to be drawn from the more socially impoverished segment of the population, i.e., from among those individuals who have the most difficulty in meeting the tasks society sets; (*b*) the diagnostic groups that were included in the hospital population showed systematic differences on these biographical variables. The schizophrenics and character disorders were similar to each other and were drawn from the most inadequate portion of the population. The manic-depressives and psychoneurotics were similar to each other and were drawn from an intermediary segment of the population, being characterized by a higher level of adaptive adequacy than that exhibited by schizophrenics and character disorders; (*c*) females in the hospital population tended to differ less strikingly from the normal population on these variables than did the males. Females were more likely to fall into the manic-depressive and psychoneurotic than into the schizophrenic and character disorder groups, and in general exhibited a higher level of premorbid social effectiveness than did hospitalized males. The relationship of adaptive adequacy to other variables has been investigated in both normal and pathological populations. It has been found to be related to prognosis in cases of schizophrenia (Phillips, 1953), success in the rehabilitation of schizophrenics (Farina & Webb, 1956), and, in normal subjects, to level of perceptual cognitive development (Lane, 1955).

The finding that a relationship exists between membership in a particular diagnostic group and the level of adaptive adequacy manifested, as measured by the biographical variables, may be related to the position recently advanced by Phillips and Rabinovitch (1958). After finding that certain symptoms tend to occur together or to be mutually exclusive in their appearance, they proposed a developmental hierarchy for categories of symptoms. Basing their proposal on conceptual grounds, Phillips and Rabinovitch (1958) suggested that symptoms indicative of Self-Deprivation and Turning against the Self, reminiscent of the manic-depressive syndrome, implied an introjection of social standards congruent with a relatively adequate level of social maturity. In contrast, symptom clusters indicative of either an Avoidance of Others or Self-Indulgence and Turning against Others, which may be taken to correspond to the schizophrenic and character disorder categories, respectively, were hypothesized as representing developmentally lower levels of psychological function. While this view is supported by the findings of Zigler and Phillips (1959), their data do not support Phillips and Rabinovitch's view that a difference in social adequacy characterizes the last two categories.

From the conceptual position advanced by Phillips and Rabinovitch and the empirical findings of Zigler and Phillips, certain predictions concerning the relationship between symptomatology and adaptive adequacy may be generated. If the biographical variables enumerated earlier are thought of as indices of adaptive adequacy, i.e., adequacy in meeting societal expectancies, then the following relationships should obtain:

---

[1] This investigation was conducted while the senior author was at the University of Missouri and the Worcester State Hospital, Worcester, Massachusetts, and was supported by the Dementia Praecox Research Project, Worcester State Hospital, and a research grant (M-896) from the National Institute of Mental Health, United States Public Health Service.

1. Individuals who manifest those symptoms conceptualized by Phillips and Rabinovitch as indicative of Turning against the Self will be more socially effective than individuals who manifest symptoms indicative of either an Avoidance of Others or Self-Indulgence and Turning against Others.
2. Since Zigler and Phillips found that hospitalized females tended to exhibit a higher level of adaptive adequacy than did hospitalized males, females should more frequently manifest those symptoms indicative of Self-Deprivation and Turning against the Self than do males. Males, on the other hand, should more frequently manifest those symptoms indicative of either an Avoidance of Others or Self-Indulgence and Turning against Others than do females.

It was the purpose here to discover if these relationships do indeed occur. The study also investigated differences in social effectiveness between individuals who manifest symptoms indicative of Avoidance of Others and those whose symptoms are indicative of Self-Indulgence and Turning against Others. Such an investigation may throw further light on the disparity between Phillips and Rabinovitch's proposal that individuals whose symptoms are indicative of Self-Indulgence and Turning against Others will evidence a higher level of adaptive adequacy than individuals whose symptoms are indicative of Avoidance of others, and the empirical findings of Zigler and Phillips which failed to provide support for such a formulation.

METHOD

The study was based on an examination of the case history data of 1,053 patients admitted to Worcester State Hospital during a 12-year period (1945–1957). This sample includes all patients referred to the Hospital Psychology Department for personality study during this period. While the sample includes a wide variety of cases, extremely deteriorated individuals, obviously organic cases, and very agitated patients are seldom referred for diagnostic appraisal and are not adequately represented in this sample. The sample was divided into a male ($N = 652$) and female ($N = 401$) group.

Categorization of symptoms. A symptom in the present study, as in the Phillips and Rabinovitch study (1958), refers to the description of a patient's behavior by a psychiatrist at the time of initial institutional contact, or the behavioral description presented by referring physicians as the primary reason for hospitalization. The symptoms include specifications, e.g., assault; general behavior patterns, e.g., irresponsible behavior; thoughts, e.g., sexual preoccupations; somatic reactions, e.g., headaches; and general affect states, e.g., being tense. Examination of the 1,053 case records resulted in the tabulation of 48 discrete presenting symptoms. Those symptoms which occurred only rarely were eliminated from further consideration because their low frequency of occurrence made them unsuitable for the type of statistical treatment employed. Consequently, the results in the male group were based on the analyses of 28 symptoms, and in the female sample, on the analysis of 24 symptoms.

A total of 29 symptoms were used. Of these, 24 had been employed by Phillips and Rabinovitch and placed by them into the categories of Avoidance of Others, Self-Indulgence and Turning against Others, and Self-Deprivation and Turning against the Self. In the earlier study, two of the symptoms reported on here, Perversions except homosexuality and Homosexuality, were treated as a single symptom, Perversions including homosexuality. Four symptoms not employed in the Phillips and Rabinovitch study were also included, Maniacal outbursts, Euphoria, Mood swings, and Fears own hostile impulses. The symptom Maniacal outbursts was placed into the Self-Indulgence and Turning against Others category. Since manic behavior has been conceptualized as a reaction against depression, the symptoms Euphoria and Mood swings were placed into the Self-Deprivation and Turning against the Self category. The symptom Fears own hostile impulses indicates an introjection of social standards in keeping with a higher level of social maturity. This symptom was therefore also included in the Self-Deprivation and Turning against the Self category.

Biographical variables. The categories within each variable were conceptualized as representing steps along a social effectiveness continuum. Because of their inherent order, the variables of age, intelligence, education, occupation, and employment history readily lent themselves to such categorization. For the variable of marital status, developmental theory as employed in recent studies (Jordan & Phillips, 1959; Lane, 1955) was utilized in ordering the categories of this variable. The categories of each variable and their order from low to high were the following:

1. Age—24 years and below; 25 to 34; 35 to 44; and 45 years and above.
2. Intelligence—IQs obtained on a standard intelligence test of 70 or less; 71 to 77; 78 to 84; 85 to 115; 116 to 124; and 125 or above.
3. Education—less than grade school; finished grade school; some high school or trade school; finished high school or trade school; some college including business college, art schools, etc.; and finished college or more.
4. Occupation—*The Dictionary of Occupational Titles* (1949) was used in placing each occupation into the categories of unskilled, semi-skilled, skilled (including foreman), service, clerical and sales, and professional and managerial. Since it was felt that placement of the categories of housewife, student, and other on a social effectiveness con-

tinuum would be arbitrary, individuals falling into these categories were not included in the analyses made on this variable.

5. Employment History—usually unemployed; seasonal, fluctuating, frequent shifts, part-time employment; and regularly employed.
6. Marital Status—single; remarried, separated, divorced, and widowed, with or without children; married without children; and married with children.

RESULTS

The first prediction to be tested was that the occurrence of particular symptoms in an individual's case history was related to that individual's social effectiveness as indicated by the biographical variables. A symptom × biographical variable contingency table for each sex was set up for each symptom on each biographical variable, e.g., tense × age, tense × intelligence, etc. Chi square was employed to determine whether the occurrence or nonoccurrence of a symptom in an individual's case history was independent of the category the individual fell into on each of the biographical variables. The .20 significance level was chosen for these tests. This modest significance level was used since the overall test of the prediction rested not upon a single finding but upon the existence of a predicted directionality in the pattern of symptom × biographical variable relationships.

The results of these analyses for men and women are presented in Tables 1 and 2, respectively. A plus indicates that the symptom was manifested more frequently by individuals who fell into the higher categories of the biographical variable, i.e., into the categories indicative of a higher level of social effectiveness. A zero indicates that the symptom was

TABLE 1

RELATIONSHIP OF SYMPTOMS TO LEVEL OF SOCIAL EFFECTIVENESS IN MEN[a]

Symptoms	*N*	Age	Marital Status	Employment History	Occupation	Education	Intelligence
Category 1—Self-Deprivation and Turning against the Self:							
Depressed	181	+	+	+	+		+
Suicidal attempt	69			+	+	+	+
Euphoria	24	+			+	+	+
Fears own hostile impulses	28		+	+		+	+
Insomnia	29		+	+		+	+
Self-depreciatory	57				+	+	+
Mood Swings	27	+				+	
Bodily complaints	86	+					
Suicidal ideas	68			+		0	+
Tense	213		+		0	+	
Headaches	39		+				0
Category 2—Self-Indulgence and Turning against Others:							
Maniacal outbursts	38		+				
Perversions (except homosexuality)	49	+		+		−	0
Drinking	148	+	0	0	−	0	0
Rape	32		+		−	−	
Robbery	66	−	−	0	−	−	0
Irresponsible behavior	33	−		0	−		
Assaultive	88	−	−		−		0
Threatens assault	71			−	−		
Emotional Outbursts	83	−	−	−			
Homosexuality	34	−	−			−	
Category 3—Avoidance of Others:							
Suspiciousness	195	+	+		+		
Perplexed	70					+	
Bizarre ideas	49		−		+		
Hallucinations	104	0	−				
Sexual preoccupations	61	−	−	0			
Apathetic	38		−				−
Withdrawn	92	−	−	−	−		−

[a] + = high level of social effectiveness; 0 = middle level of social effectiveness; − = low level of social effectiveness.

TABLE 2

RELATIONSHIP OF SYMPTOMS TO LEVEL OF SOCIAL EFFECTIVENESS IN WOMEN[a]

Symptoms	N	Age	Marital Status	Employment History	Occupation	Education	Intelligence
Category 1—Self-Deprivation and Turning against the Self:							
Suicidal attempt	69				+	+	+
Suicidal ideas	68					+	+
Euphoria	25			+		+	
Doesn't eat	28	+		+			
Bodily complaints	60	+					
Tense	138	+			0	+	+
Headaches	23	+	0				
Self-depreciatory	59	0	+	+	+	+	0
Depressed	167	+	+		0		0
Fears own hostile impulses	21	−	+				
Mood swings	27	−					
Category 2—Self-Indulgence and Turning against Others:							
Drinking	33	0	0	+			
Assaultive	37	−				−	
Maniacal outbursts	49	−	−	−			
Threatens assault	23	−	−	−	−		
Irresponsible behavior	36		−		−	−	−
Emotional outbursts	60	−	−	−	−	−	
Category 3—Avoidance of Others:							
Perplexed	72	−	+		+	+	0
Hallucinations	68	−		+	+		
Bizarre Ideas	34	0		−	+		
Sexual preoccupations	33	0		0			
Apathetic	38			−	−	0	
Withdrawn	53		−				
Suspiciousness	132		−				−

[a] See footnote Table 1.

manifested more frequently by individuals who fell into the middle categories on the biographical variable. A minus indicates that the symptom was manifested more often by individuals who fell into the lower categories of the biographical variables. No mark indicates that no significant relationship was found between the occurrence of a symptom and the categories on the biographical variable. While symptoms were grouped into the original symptom categories employed by Phillips and Rabinovitch, within categories the symptoms were listed from high to low, i.e., from the most pluses to the most minuses. A summary of the findings reported in Tables 1 and 2 is presented in Table 3.

Symptom Category 1 vs. Symptom Categories 2 and 3. As can be seen in Table 3, the findings support the prediction that individuals who manifest those symptoms indicative of Self-deprivation and Turning against the Self are more socially effective than individuals who manifest symptoms indicative of either Avoidance of Others or Self-Indulgence and Turning against Others. For men, the analyses done on symptoms in Category 1 (Turning against the Self) disclosed that in 32 of the 35 instances in which a significant relationship was found between the occurrence of a symptom and a biographical variable, the symptom tended to occur in individuals who fell into the more socially effective categories on the biographical variables. Among symptoms in Categories 2 (Turning against Others) and 3 (Avoidance of Others), in only 10 of 53 instances in which significant relationships were found did the relationships indicate that the symptom tended to occur more frequently in more socially effective individuals. This general pattern was also found in women. For women, the analyses done on symptoms in Category 1 revealed that in 21 of the 29 instances in which a significant relationship was found between the occurrence of a symptom and a biographical

TABLE 3

SUMMARY OF RELATIONSHIPS BETWEEN SYMPTOM CATEGORIES AND SOCIAL EFFECTIVENESS IN BOTH MEN AND WOMEN[a]

Symptom Category	Men			Women		
	No. & % +	No. & % 0	No. & % −	No. & % +	No. & % 0	No. & % −
Category 1 Self-Deprivation and Turning against the Self	32 (91%)	3 (9%)	0 (0%)	21 (72%)	6 (21%)	2 (7%)
Category 2 Self-Indulgence and Turning against Others	5 (14%)	9 (26%)	21 (60%)	1 (5%)	2 (10%)	18 (85%)
Category 3 Avoidance of Others	5 (28%)	2 (11%)	11 (61%)	6 (32%)	5 (26%)	8 (42%)

[a] See footnote Table 1.

variable, the relationships indicated that the symptom tended to occur more frequently in more socially effective individuals. This was true in only 7 of the 40 instances in which significant relationships were found between symptoms and biographical variables in Categories 2 and 3.

Symptom Category 2 vs. Symptom Category 3. As can be seen in Table 3, there was a tendency in the male group for symptoms indicative of an Avoidance of Others to be related to a higher degree of social effectiveness than was true for symptoms indicative of Self-Indulgence and Turning against Others. This tendency was even more marked in the female group.

Category of Symptom × Sex. The second prediction was that females more frequently manifest those symptoms indicative of Self-Deprivation and Turning against the Self than do males, while males more frequently manifest symptoms of either an Avoidance of Others or Self-Indulgence and Turning against Others. A symptom × sex contingency table was set up for each symptom. Chi square was used to estimate whether the occurrence or non-occurrence of a symptom in an individual's case history was independent of that individual's sex. Since this prediction was based on specific symptom × sex relationships, the .05 level of significance was chosen for this group of tests. As can be seen in Table 4, the findings support the prediction made. Of the 7 symptoms in the Self-Deprivation and Turning against the Self category whose occurrence was found to be significantly related to the

TABLE 4

RELATIVE INCIDENCE OF SYMPTOM OCCURRENCE FOR MEN AND WOMEN

Symptom	Occurs More Frequently In
Category 1— Self-Deprivation and Turning against the Self:	
Suicidal attempt	Women
Suicidal ideas	Women
Euphoria	Women
Doesn't eat	Women
Bodily complaints	*ns**
Tense	*ns*
Headaches	*ns*
Self-depreciatory	Women
Depressed	Women
Fears own hostile impulses	*ns*
Mood swings	Women
Insomnia	*ns*
Category 2— Self-indulgence and Turning against Others:	
Maniacal outbursts	Women
Perversions (except homosexuality)	Men
Drinking	Men
Rape	Men
Robbery	Men
Irresponsible behavior	Women
Assaultive	Men
Threatens assault	Men
Emotional outbursts	*ns*
Homosexuality	Men
Category 3— Avoidance of Others:	
Suspiciousness	*ns*
Perplexed	Women
Bizarre ideas	*ns*
Hallucinations	*ns*
Sexual preoccupation	*ns*
Apathetic	Women
Withdrawn	*ns*

* *ns* (not significant).

sex variable, all tended to occur more frequently in women than in men. In the other two symptom categories, of the 11 symptoms whose occurrence was found to be significantly related to sex, 7 tended to occur more frequently in men than in women, while 4 tended to occur more frequently in women than in men. It should be noted that it is the symptoms in the Self-Indulgence and Turning against Others category that tend to occur more frequently in men than in women. This finding does not hold for symptoms in the Avoidance of Others category. Of the two symptoms in the latter category whose occurrence was found to be significantly related to sex, both tended to occur more frequently in women.

DISCUSSION

The most striking finding of the present study is that a systematic relationship exists between measures of social effectiveness and particular groups of symptoms. This observation lends considerable support to the view of Phillips and Rabinovitch (1958) that symptoms are indices of an individual's basic orientation towards the self and others and that this orientation is in turn related to the developmental level the individual has attained. The hypothesis that the most socially effective among hospitalized persons manifest symptoms indicative of Self-Deprivation and Turning against the Self is clearly borne out. Individuals manifesting symptoms of either Self-Indulgence and Turning against Others or Avoidance of Others tend to be drawn from a less socially effective segment of the population.

Another deduction from the major hypothesis was that because hospitalized women appear to be drawn from a more socially effective portion of the population than hospitalized men (Zigler & Phillips, 1959), women more frequently manifest higher order symptoms than men. This prediction is supported by the findings. In every case where a significant relationship existed between sex of the individual and occurrence of a symptom in the Self-Deprivation and Turning against the Self category, it was found that the symptom tended to occur more frequently in women. On the other hand, where significant relationships were found between sex and symptoms of the other two categories combined, a majority of the symptoms tended to occur more frequently in men.

There is no support here for Phillips and Rabinovitch's conceptualization that the Self-Indulgence and Turning against Others category is indicative of a higher level of psychological functioning than is the Avoidance of Others category. In fact, the findings suggest that individuals who manifest symptoms in the latter category tend to be more socially effective than those who manifest symptoms in the former category. This tendency of the Avoidance of Others category to be associated with a higher level of social effectiveness than the Turning against Others category is especially marked in women. Further evidence that the former category represents a higher level of functioning than the latter is indirectly available in the finding that in the two instances where a significant relationship existed between sex and the occurrence of symptoms in the Avoidance of Others category, the symptoms tended to occur more frequently in women. Since the highest order symptoms, i.e., those in the Turning against Self category, tended to occur more frequently in women, this suggests that the Avoidance of Others category shares features in common with this category. Contrary to the common relationship between these two categories, symptoms in the Turning against Others category predominantly tended to occur more frequently in men than in women.

The finding that the Avoidance of Others category is associated with a higher level of social effectiveness than the Turning against Others category presents certain theoretical difficulties. To the extent that Avoidance of Others and Turning against Others are related to the schizophrenias and the character disorders, respectively, it follows that schizophrenia is a "higher" level disturbance than the character disorders. This view would run contrary to a generally held position that schizophrenia represents a dominance of what may be genetically considered the most primitive type of functioning, whereas the character disorders represent a somewhat higher level of psychological activity. (Fenichel, 1945; Friedman, 1952; Osborne, 1940). This somewhat paradoxical finding challenges explanation.

If, however, one examines those symptoms within the Avoidance of Others category which are regularly associated with the schizophrenic syndrome, such as Withdrawn, it will be observed that these tend to occur in less adequate individuals. It is the symptoms more atypical of this psychotic disorder which occur in more socially adequate patients. If it is then assumed that the Avoidance of Others category is indeed representative of the schizophrenias, the findings of the present study would not demand the assertion that the overt schizophrenic functions more effectively than the individual hospitalized with a character disorder. What has been measured in the present study is the premorbid social effectiveness of individuals. It may be that the schizophrenic actually achieves a higher level of social effectiveness than does the character disorder, but with onset of the illness, he manifests a more encompassing regression to genetically lower levels of function. This view leads to a conception of schizophrenia as a process in which genetically early levels of functioning infringe upon the higher, more recently developed genetic characteristics. On the other hand, the character disorders would be viewed as a failure in reaching higher levels of development.

Finally, social effectiveness has been treated in the present study as if it were a unitary phenomenon. Actually, it would appear more reasonable to view adaptive adequacy as a constellation of a number of relatively independent patterns of behavior. This view is supported by the recent findings (Jordan & Phillips, 1959) that most items on a frequently used social effectiveness scale, the Worcester Scale of Social Attainment (Phillips & Cowitz, 1953), fell into two major clusters. One cluster appeared to represent a factor of social participation; the other was related to striving for personal advancement. In the present study, the biographical items employed to represent the continuum of social effectiveness appear to be more congruent with the "striving" than with the "social participation" dimension. This appears to be true most explicitly for the occupation, education, and employment history variables. The dimension of "social participation," on the other hand, seems to find minimal representation in the case history items employed. Had biographical items more representative of degree of social participation been used, then individuals whose symptoms are indicative of Avoidance of Others might well have scored at a lower level of adaptive adequacy.

The finding that women tend toward self-destructive behaviors, while men manifest either self-indulgence or behaviors injurious to others, is consistent with the earlier finding that hospitalized women tend to be more effectively adaptive in society than hospitalized men (Zigler & Phillips, 1959). The question arises as to why this should be the case. Tentatively, the suggestion can be offered that both social effectiveness and pattern of symptomatic behavior represent the degree to which an individual tends either to resist or to conform to the mores of society. Thus, men appear to be more aggressive and nonconformers. Other studies (Brun-Gulbrandsen, 1958; Child, Potter, & Levine, 1946) have indicated that women are subjected to far greater pressure for social conformity than are men, and that any deviation in their behavior, particularly in the direction of aggression, is looked upon with much greater concern than it is in men.

The present study has dealt with the proposition that symptom clusters are differentially related to the adaptive adequacy of the individual. Two additional propositions have been advanced in this and earlier papers (Phillips & Rabinovitch, 1958). First, the symptom clusters are indicative of one's role orientation in society; second, the symptoms which one manifests, as well as the congruent level of adaptive adequacy, reflect the psychological development or personal maturity which one has achieved. However, the studies so far reported have dealt only with psychopathological populations. It has been established that either Avoidance of Others or Self-Indulgence or Turning against Others characterize socially inadequate individuals, whereas those psychiatric cases who Turn against the Self tend to be drawn from a more effective stratum of society. It is difficult to conceptualize such negatively toned role orientations as typifying normal individuals drawn from equivalent segments of society. This raises the interesting question of whether consistent patterns of social interaction may be found in a normal population at different levels of social adequacy and whether these orientations are in any way related to those of pathological individuals of the same levels of social effectiveness.

SUMMARY

The present study tested the following predictions:

1. Hospitalized individuals who manifest those symptoms indicative of Turning against the Self have a more socially effective premorbid history than those with symptoms of either Avoidance of Others or Self-Indulgence and Turning against Others.

2. Hospitalized females more frequently manifest symptoms of Self-Deprivation and Turning against the Self than do males. Males, on the other hand, more frequently manifest symptoms of either Avoidance of Others or Self-Indulgence and Turning against Others than do females.

The study also investigated differences in social effectiveness between individuals who manifest symptoms of Avoidance of Others and those whose symptoms are indicative of Self-Indulgence and Turning against Others. The variables of age, intelligence, education, occupation, employment history, and marital status were employed as indices of social effectiveness. The study was based on an examination of the case history data of 1,053 patients (652 men and 401 women).

Both predictions were supported by the findings. The results also suggested that individuals whose symptoms are indicative of Self-Indulgence and Turning against Others tend to be less socially effective than individuals whose symptoms are indicative of an Avoidance of Others. Present findings were discussed in relation to those of earlier studies as were their implications for patterns of social interaction in a normal population.

REFERENCES

BRUN-GULBRANDSEN, S. Kjnnsrolle og ungdomskriminalitet. Oslo: Universitetsforlaget, 1958.

CHILD, I., POTTER, E., & LEVINE, E. Children's textbooks and personality development: An exploration in the social psychology of education. *Psychol. Monogr.*, 1946, **60**(3, Whole No. 279).

FARINA, A., & WEBB, W. Premorbid adjustment and subsequent discharge. *J. nerv. ment. Dis.*, 1956, **124**, 612–613.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FRIEDMAN, H. Perceptual regression in schizophrenia: An hypothesis suggested by the use of the Rorschach test. *J. genet. Psychol.*, 1952, **81**, 63–98.

JORDAN, E., & PHILLIPS, L. Dimensions of social maturity. Unpublished manuscript, Worcester State Hospital, 1959.

LANE, J. Social effectiveness and developmental level. *J. Pers.*, 1955, **23**, 274–284.

OSBORNE, R. Palaeophrenia: A re-evaluation of the concept of schizophrenia. *J. ment. Sci.*, 1940, **86**, 1078–1085.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

PHILLIPS, L., & COWITZ, B. Social attainment and reactions to stress. *J. Pers.*, 1953, **22**, 270–283.

PHILLIPS, L., & RABINOVITCH, M. Social role and patterns of symptomatic behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 181–186.

UNITED STATES GOVERNMENT PRINTING OFFICE. *The dictionary of occupational titles.* Washington: USGPO, 1949.

ZIGLER, E., & PHILLIPS, L. Case history data and psychiatric diagnosis. Unpublished manuscript, Worcester State Hospital, 1959.

(Received July 28, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 239-245

ROLE ASSIGNMENT AND INTERPERSONAL INFLUENCE[1]

IVAN D. STEINER AND WILLIAM L. FIELD
University of Illinois

LABORATORY investigations of small groups have usually involved the study of contrived groups consisting of strangers who are assembled for a brief period of time and who operate with full knowledge that their associations with one another will not persist into the future. Such contrived groups are manifestly unlike the real-life groups to which we often like to generalize our research findings. In addition to lacking a history and a future, they ordinarily lack a role system that is appropriate to the task at hand. Members are not assigned, nor need they assume, specialized functions; expectations that specific members have special abilities, loyalties, obligations, or responsibilities are seldom created, and the contrived group functions as an unstructured aggregation of individuals.

When investigators have assigned roles to the individuals studied in laboratory experiments, the consequences of such assignments have been very apparent. Carter, Haythorn, Shriver, and Lanzetta (1950) found that assignment of the leadership role to one of the members of a group resulted in leader-follower interactions that were different from those which occurred when the leader was permitted to "emerge" without assignment. Smith (1956) demonstrated that the effect of a silent partner on the productivity and satisfaction of group members depended upon whether or not the group had a role system in which the partner's silence was appropriate. Berlo (1956) assigned each member of his experimental groups the responsibility for performing specific procedural functions, and found that members of such groups were more productive and more satisfied with their performance than were members of similar groups in which no role assignments were made. Borgatta and Bales (1953) have reported that groups in a role playing situation produced interaction profiles which were different from those produced by groups operating without assigned roles. And Steiner and Dodge (1956) discovered that accurate interpersonal perceptions were responsible for increased individual and group productivity when members had not been assigned roles; when the group operated according to a role system, accuracy of interpersonal perceptions had almost no bearing upon individual and group productivity.

The study reported by this paper is like the Steiner and Dodge investigation in that it is concerned with the indirect, rather than the direct, consequences of role assignment. It explores the conjunctive or interactional effects that occur when role assignments are made, and it attempts to demonstrate that persons' perceptions of one another, and their reactions to one another, are contingent upon the presence of a role system. It is hoped that this line of investigation may provide insight concerning the probable generality of conclusions generated by laboratory studies of conformity, attitude change, and group dynamics.

THEORY AND HYPOTHESES

Roles are sometimes "imposed from above" with a minimum of concern for the preferences of the persons involved. Role assignments in military organizations often have this arbitrary quality. Somewhat less arbitrary are the role assignments that occur when persons are appointed or elected to "offices" by their associates; here it is reasonable to presume that the associates have given some consideration to the attitudes and personality of the individual selected. Still less arbitrary are those role assignments that occur when role systems develop slowly, and persons have considerable freedom to "move into" roles that harmonize with their attitudes and personalities. The gradually emerging role structure of a growing business concern is likely to permit this type of self-assignment.

---

[1] This investigation was supported by a research grant (M-1774) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service (F. E. Fiedler, Principal Investigator). The writers are indebted to Paul Pollock for a superb performance as the experimenter's accomplice, and to William McGuire and Alan Bass for criticisms of the original manuscript.

Two postulates that are concerned with these three methods of role assignment, and with the condition in which no role assignments at all are made, underlie the hypotheses tested by this research.

Postulate 1. The propensity to interpret behaviors as indicative of personal preferences is greatest when no roles are assigned, less when there is self-assignment to roles, still less when assignment is by appointment or election, and least when roles are imposed from above.

The fact that individuals are aware of the discrepancy between personal preferences and the behaviors required by roles imposed from above is indicated by comments such as: "It's not his fault; he's just doing what he has to do," or "The Captain is pretty GI when he's on duty, but he's probably a good Joe when he has a chance." The frequency with which such statements are made by military personnel, employees of large business firms, and others who operate within role-organized agencies suggests that people regularly "make allowances" for role restrictions when estimating others' preferences, wishes, or personality. Postulate 1 asserts that this tendency to make allowances is greater when roles are assigned than when they are not, and greater when roles are assigned arbitrarily than when there is a high degree of self-assignment.

The second postulate contends that the person becomes aware of his tendency, and the tendency of his associates, to make allowances. Consequently, he does not expect that his associates will construe his role-determined behaviors as indicating his personal preferences. Thus, the "civilian soldier" may commit numerous acts of brutality during wartime without feeling that important reference persons will regard him as a brutal person. However, the individual's belief that his behaviors will not be interpreted as indicating personal preferences is undoubtedly most appropriate and most easily maintained when roles are imposed from above; it should be less easily maintained when role assignments are made by election or self-selection.

Postulate 2. The propensity to expect others to construe one's behavior as indicating personal preferences is greatest when no roles are assigned, less when there is self-assignment to roles, still less when assigment is by appointment or election, and least when roles are imposed from above.

Although the postulates deal with four degrees or positions along a presumed continuum, the research to be reported below is concerned with only the two most extreme conditions: no role-assignment vs. role assignment from above. Hypotheses are thus stated in terms which deal only with these two extreme conditions, even though relationships involving the less extreme conditions are clearly implied.

Because the individual is presumed (Postulate 1) to accept observable behaviors as trustworthy clues to an associate's preferences when no roles are assigned, but not when roles are assigned from above:

Hypothesis 1. When no role assignments are made, persons feel more confident of their ability to evaluate the personal preferences of an individual who produces deviant behavior than is the case when roles are imposed from above.

Hypothesis 2. When an imposed role makes it appropriate for a person to produce unpopular and deviant behavior, he is less severely rejected for such behavior than is the case when he produces such behavior without the role having been imposed.

Hypothesis 3. When an imposed role makes it appropriate for a person to produce popular behavior, performance of such behavior leads to a smaller increase in his popularity than is the case when he performs the behavior without the role having been imposed.

According to Postulate 2, an imposed role should decrease the person's concern about whether or not his behaviors will lead to rejection. Consequently, he need not be so alert to observe the reactions that his behaviors are evoking in others; he can afford to persist in a course of action without permitting or inviting others to react to what he is doing. If the activities of the group involve the discussion of a controversial issue, he can present his arguments at great length without allowing others to react to them. At the same time, his associates will realize that these views do not necessarily represent his true feelings, and will feel less need to interrupt in order to correct his "erroneous thoughts." Consequently, if a communication be defined as any continuous speech behavior produced by a single person, we may hypothesize:

Hypothesis 4. When roles are imposed, members of a discussion group produce longer communications than is the case when no roles are imposed.

Because imposed roles make it easy for the person to conclude that the behaviors of his associates do not represent their true preferences or opinions, those behaviors should have little impact upon his own preferences and opinions. "They don't even believe what they are saying, so why should I?" Consequently:

Hypothesis 5. When an imposed role makes it appropriate for a person to express deviant opinions, his behavior has less effect on the opinions of his associates than is the case when those deviant opinions are expressed without a role having been imposed.

METHOD

Preliminary Session

Ninety-five male college students enrolled in the elementary psychology course at the University of Illinois participated in this study. They received laboratory credit for their participation. During the first session they were assembled in a large classroom and told that they were to serve as subjects (*S*s) in a study of word meanings. *S*s were given a three-page instrument on which they were asked to rate the meanings of three concepts: mental illness, desegregation of public schools, and going steady. Each page presented the same ten graphic rating scales, the ends of which were identified by polar adjectives such as wise-foolish, worthless-valuable, and good-bad. Both the seven-step rating scales and the polar adjectives were taken from Osgood and Suci's semantic differential (1955). As is customary with this kind of instrument, *S*s indicated their appraisals of the three concepts by placing check marks on the scales. Only their appraisals of desegregation of public schools were scored; ratings of the other concepts were obtained to lessen the possibility of sensitizing the *S*s to the desegregation issue. Furthermore, responses to only six of the ten rating scales were considered in computing the *S*'s attitude toward desegregation of schools; these were six scales that Osgood had found to load heavily on the "evaluative" dimension. Four other scales dealt with potency and activity, and were not scored. Responses at the favorable end of a scale were assigned a value of seven, while responses at the unfavorable end were assigned a value of one. With six items, scores could range from 6 to 42; the obtained scores ranged from 7 to 42, with a mean of 32.5.

Laboratory Session

Subjects. Sixty-eight of the *S*s who participated in the preliminary session and whose scores on the rating scales indicated that they held nonsegregationist attitudes were selected to participate in the next phase of the research. Their scores ranged from 24 (the theoretically neutral score) to 42 (the most prodesegregationist score possible). Eighty-one of the original 95 *S*s had scores of 24 or larger, but 3 were discarded because they were Negroes, and scheduling difficulties prevented the use of 10 others.

Each of the 68 *S*s was paired with another *S* whose attitude score was approximately the same as his own. The two *S*s in each pair were scheduled to arrive at the laboratory at the same time. Upon arrival the pair was joined by another student who was serving as an accomplice of the experimenter (*E*). The same individual served as accomplice in all of the 34 experimental groups.

The task. The accomplice and the two naive *S*s were seated at a table, the accomplice always occupying the position at *E*'s left, while one naive *S* sat opposite *E*, and the other sat at *E*'s right. A cardboard sign bearing the letter A was placed in front of the accomplice; a sign bearing the letter B was placed in front of the naive *S* who sat opposite *E*, and a sign bearing the letter C was placed in front of the *S* at *E*'s right. These letters substituted for names throughout the experimental session.

E explained to the *S*s that their task was to discuss the desirability of desegregation of public schools, and that they should attempt to reach agreement among themselves during the 15-min. period allowed for discussion. *E* then said:

> In order that your discussion will not be one-sided, you are urged to take account of the points of view held by a typical Southern segregationist; a typical Northern minister, priest, or rabbi; and a typical member of the National Association for the Advancement of Colored People.

This admonition was also written on paper and placed before each *S* where it could serve as a reminder during the discussion.

Role assignments. At this point, *S*s in half of the 34 groups received role assignments. *E* said:

> Now, so that we will be sure that each of these three points of view is adequately presented in your discussion, I'm going to assign each of you the responsibility for presenting one of these three points of view.

E placed a sign in front of Person A (the accomplice) which said Southern Segregationist; signs placed in front of Persons B and C said Northern Clergyman and NAACP Member, respectively. *E* then instructed each *S* to pretend that he was the kind of person designated by his sign, and to advance the views such a person would advance.[2]

---

[2] The role of Northern Clergyman and NAACP Member were selected for use in this study because it was expected that *S*s would conceive such roles to harmonize with their own nonsegregationist attitudes, and would not feel that their role restricted their freedom to express their own opinions. Apparently this expectation was correct. *S*s sometimes prefaced their initial comments by saying, "Because I'm a Northern Clergyman (or NAACP Member) I think desegregation

Groups selected to receive the role-assignment instructions were matched with groups not receiving such instructions on the basis of attitude scores obtained from the preliminary session. *S*s receiving the two types of treatment had identical mean scores of 33.9; the standard deviations of the scores were 4.6 for the role-assigned groups and 4.0 for the no-role groups.

Behavior of the accomplice. The accomplice was always assigned the role of the segregationist in the role groups, and when no roles were assigned he expressed the same segregationist views. He was coached before the experimental sessions began to maintain the position that, although Negroes might not be inherently inferior, their inferior social and economic backgrounds made desegregation unwise during any time in the foreseeable future. Pretest groups were run to give the accomplice practice in presenting standard arguments, and to enable the accomplice and *E* to anticipate the arguments which would be advanced by naive *S*s during the laboratory sessions. In this way the accomplice was equipped with a standard reply for most of the statements made by the naive *S*s. Unanticipated arguments were either ignored, or were answered by the accomplice in a way which seemed consistent with his prescribed segregationist viewpoint.

Certain other restrictions were placed upon the accomplice. He was to attempt to speak first, since the pretests had indicated that it was easier for him to outline his general segregationist viewpoint if he spoke before others had expressed their contrary views. In all but one of the 34 groups he was successful in his attempt to speak first. Furthermore, he was to avoid speaking more often, or for a longer time, in one type of group than in the other. Data to be presented later indicate that he was also successful in meeting this requirement. Finally, even though his "partners" might become emotional during the discussion, he was forbidden to use sarcastic or derogatory language, or to interrupt them while they were speaking. The investigator as *E* observed that these rules were strictly observed.

Measurements. While the discussion was in progress, *E* employed a polygraph to produce a record of the length and serial position of each *S*'s contributions. In order to identify each speaker, three event markers were employed. Whenever a person spoke, *E* pressed a button that activated the appropriate event marker; the button was kept in the depressed position until the *S* finished speaking or was interrupted by another *S*.

At the conclusion of the discussion period, *S*s expressed their attitudes toward desegregation of schools on the same semantic differential scales as had been employed in the preliminary session. They also made other ratings to be described later in this paper.

Explanation to Ss. Before *S*s were dismissed they were given a brief but truthful explanation of the study. Reasons why the accomplice had taken a segregationist stand were revealed, and *S*s were assured that neither the accomplice nor the *E* was a segregationist or interested in convincing them that they should be. Finally, *S*s were requested not to discuss the study with anyone before the semester ended. So far as can be determined, this request was uniformly honored by the *S*s.

RESULTS

Hypothesis 1

Immediately after the discussion period, *S*s were asked to rate their partners' attitudes toward desegregation on the same 10-item scale as was used in the preliminary session. When these ratings had been made, the *S*s were asked to indicate their confidence in the ratings they had made of each of their partners by placing a check mark on a seven-step graphic rating scale. *E* subsequently assigned numerical values to these check marks which ranged from 1 for a check mark in the space labeled "sheer guesses" to 7 for a check mark in the space at the other end of the scale labeled "very accurate." The mean of the scores assigned to ratings of the accomplice made by *S*s in the role-assigned groups was 4.3, while the mean for the ratings made by *S*s in the no-role groups was 5.0. A four-celled chi square test, with *S*s divided into a high and a low scoring category, yielded a chi square of 5.22, with a probability of less than .03. In computing this chi square, *S*s were divided as closely as possible at the median score (between scores of four and five), and Yate's correction was applied.

The mean of the scores assigned to ratings of nonaccomplice partners in role-assigned groups was 4.4, while in no-role groups it was 4.6. This difference was not statistically significant.

Hypotheses 2 and 3

Rejection of the accomplice and acceptance of the naive partner were evaluated by five sociometric questions administered at the conclusion of the discussion. *S* was asked which of his two partners he would prefer as a roommate, which would prefer him as a roommate, which had the better understanding of him, which he would choose to accompany him on a week-end camping trip, and with which he would be more inclined to discuss his personal problems. Interrelationships between answers to these five questions were determined by phi coefficients to range from +.28 to +.67. All of the interrelationships were significant at the

---

of schools is a good idea." After this initial acknowledgment of the role, *E* was unable to observe any difference between the behaviors of *S*s in the role-assigned and no-role groups.

.02 level or better. Consequently, the five items were pooled to form a sociometric preference scale yielding scores ranging from zero to five. (The score is the number of times the *S* chose the accomplice rather than his naive partner.) The distribution of scores was highly skewed, with 27 *S*s receiving the score of zero because they had not chosen the accomplice in response to any of the five items. For this reason a four-celled chi square test would have been relatively insensitive, and a six-celled test was employed instead. With scores of five, four, and three collapsed to form one category, and scores of two and one collapsed to form another, the obtained chi square value was 7.08 which, with two degrees of freedom, represents a probability of less than .05.

Although this finding indicates that the discrepancy between the sociometric acceptance accorded the accomplice and the naive *S* is greater in the no-role groups, the data do not tell us whether this outcome represents a confirmation of Hypothesis 2, of Hypothesis 3, or of both. Information on this point was obtained from additional postdiscussion ratings. *S*s were given a series of 18 graphic rating scales on which to rate their two partners. Each scale had six steps, and the ends of each scale were identified by polar adjectives of an evaluative nature. Examples are cooperative–uncooperative, and friendly–unfriendly. Ratings at the good ends of the scales were scored six, and ratings at the bad end were scored one. Numerical values were summed across all 18 scales to obtain the *S*'s esteem for a partner. The mean of these esteem ratings received by naive *S*s was 84.4 in the no-role groups and 80.7 in the role-assigned groups. The difference is in the direction predicted by Hypothesis 3, and is significant at the .10 level by a one-tailed *t* test. The mean of the esteem ratings given to the accomplice was 78.3 in the no-role groups and 78.8 in the role-assigned groups. Although this difference is in the direction predicted by Hypothesis 2, it is extremely small and does not approach statistical significance. Apparently the tendency for *S*s in the no-role groups to show a greater sociometric preference for the naive partner than do *S*s in role-assigned groups reflects their relatively higher esteem for the naive partner, rather than lower esteem for the accomplice. It should be noted, however, that in both types of groups the naive partner received significantly more sociometric choices and significantly higher esteem ratings than did the accomplice.

Hypothesis 4

The polygraph records indicated that the average duration of the communications produced by the naive *S*s in the role-assigned groups was 31.5 sec. In the no-role groups the average duration of communications was 23.8 sec. To test the significance of the difference between types of groups, the mean duration of each *S*'s communications was computed, and the two distributions of means were tested by the Mann-Whitney *U* test. The *U* test was employed for this purpose because the distributions were highly skewed. In accordance with Hypothesis 4, the distribution for the role-assigned *S*s was found to be stochastically larger than the distribution for no-role *S*s, the difference being significant at the .04 level by one-tailed test.

The accomplice, whose job it was to maintain the same pattern of behavior in both types of groups, produced communications in the role-assigned groups that averaged 43.5 sec. in length. In the no-role groups his communications averaged 41.0 sec. in length. In the no-role groups his total speaking time averaged 460 sec. per group, while in the role-assigned groups the average was 478 sec. Neither of these differences even approaches statistical significance. Apparently the accomplice was able to abide by his instructions rather well.

Hypothesis 5

The impact of the accomplice on the attitude of the naive *S*s was measured by the discrepancy between *S*s' pre- and postdiscussion attitude scores. The second of these scores was subtracted from the first. The mean difference was +4.6 for *S*s in the no-role groups, and +.7 for *S*s in the role-assigned groups. This outcome is in the predicted direction (large positive difference between pre- and postdiscussion scores represents large change in the direction advocated by the accomplice), and is significant at the .01 level by a one-tailed *t* test.

DISCUSSION

This investigation has employed a univariate design in which numerous variables have been held constant while one variable (role assignment vs. no role assignment) has been manipulated. Whenever this type of design is used there is the possibility that relationships between the independent and dependent variables are contingent upon the level at which other variables have been held constant. Consequently, it is highly desirable that the present study be repeated, and that systematic changes be made in the levels at which variables are controlled. It is especially important that the identity of the accomplice be varied in order to determine whether his personal characteristics and his conception of his role influence the outcome of the research. Other variables that should be subjected to systematic investigation are: the issue that is discussed; the size and direction of the discrepancies between the positions held by the accomplice and the naive *S*s, and the personal characteristics of the naive *S*s. Until such replications have been performed, it can only be said that the hypotheses have been tested under conditions that appear not to be highly atypical, and that they have received fairly strong support.

Although the data support four of the five hypotheses, *S*s were not found to reject the accomplice more completely in the no-role than in the role-assigned groups. Furthermore, *S*s in no-role groups did not judge the accomplice to be more segregationist in attitude than did *S*s in role-assigned groups; both kinds of *S*s judged him to be a strong segregationist. These findings appear to contradict the line of reasoning on which this research has been based, and make it difficult to understand why *S*s in the no-role groups produced shorter communications, yielded more completely to the segregationist view advocated by the accomplice, and esteemed their naive partner more highly than did *S*s in the role-assigned groups. Perhaps a reasonable explanation is that *S*s in the no-role groups were more certain of their appraisals of the accomplice (Hypothesis 1), and that a difference in certainty of appraisal has much the same effect as does a difference in the content of the appraisal itself. However, if this is true, it is still unclear why *S*s in the no-role groups did not indicate less esteem for the accomplice.

Perhaps part of the difficulty lies in the fact that data were obtained from rating scales. College students are undoubtedly reluctant to rate one another unfavorably, and the instrument with which esteem data were obtained permitted them to avoid doing so. On the other hand, the sociometric data were obtained from forced-choice items, and here the accomplice was rated unfavorably by both types of *S*s, and especially so by the *S*s in no-role groups.

Some of *E*'s observations are also pertinent at this point. Because the accomplice was a very persuasive person, and because the research design required that he behave the same way in both types of groups, his performance in the role-assigned groups appeared to be very genuine, and to reflect his true feelings. Because of the enthusiasm with which he expressed segregationist views, many naive *S*s may have "forgotten" that he had been assigned the role of a segregationist, or may have concluded that his assigned role happened to harmonize with his attitudes. If such was the case, some of the effect of role assignment was undoubtedly nullified by his great ability to enact the role. In this respect we undoubtedly failed to approximate the conditions in real-life groups where the individual who is assigned a role to perform is unlikely to be as seemingly sincere or strenuous an advocate as was our accomplice.

An alternative explanation of the data cited in support of Hypothesis 5 must be considered. Janis and King (1954) have found that *S*s who are assigned roles requiring that they advocate a point of view sometimes shift their attitudes to favor the view they have been induced to advocate. Consequently, the failure of our role-assigned *S*s to yield to the accomplice's arguments might be a result of their having played desegregationist roles, rather than a result of their having doubted the sincerity or genuineness of the accomplice's arguments. If this alternative explanation of our data is correct, *S*s assigned the most extreme desegregationist role (NAACP Member) and *S*s who spoke most often and at greatest length while enacting their assigned role, should show the most resistance to the accomplice's arguments. None of these derivations is confirmed; the obtained probabilities are greater than .70. It therefore seems appropriate to conclude that the findings cannot be explained as consequences of the naive *S*s' enactment of their own roles.

The present study supplements the earlier investigation by Steiner and Dodge (1956). Whereas the earlier study was concerned with the manipulation of interpersonal perceptions, and with the consequences of such manipulations on sociometric preferences and group productivity, the present study has permitted interpersonal perceptions to vary in response to the assignment of roles, and has explored some presumed consequences of these variations in interpersonal perceptions. As in the case of the earlier study, the presence or absence of role assignments has been found to influence the relationship between social psychological variables. Behavior that evokes one pattern of response in a group operating within the framework of a role system has been found to evoke a quantitatively different pattern of response when no roles are assigned. It seems reasonable to conclude that role assignment "interacts" with other variables to produce effects that are very important both to the student of group dynamics and to the practitioner who would apply the scholar's wisdom outside the walls of the laboratory. The study suggests that very great care must be taken when generalizing the findings of laboratory investigations concerned with conformity, attitude change, or group discussion.

SUMMARY

Ninety-five male college students responded to an instrument designed to evaluate their attitudes toward desegregation of schools. From this number, 68 whose attitude scores reflected nonsegregationist views were selected to participate in a laboratory study. Thirty-four groups, each consisting of two *S*s drawn from this sample and another college student who served as an accomplice of the *E*, discussed the desegregation issue for a period of fifteen minutes. In half of the groups the *S*s were assigned roles; the accomplice was always assigned the role of a Southern segregationist while the other participants were assigned the role of a Northern clergyman and the role of a member of the NAACP. In the other half of the groups, participants were not assigned specific roles but were asked to make certain that the views of these three kinds of persons were presented in their discussion. The accomplice endeavored to express the same segregationist views in all groups.

When specific roles were not assigned, the nonaccomplice *S*s: (*a*) had greater confidence in their appraisals of the accomplice's attitudes, (*b*) indicated greater sociometric preference for one another (rather than for the accomplice) at the conclusion of the discussion period, (*c*) produced shorter communications, and (*d*) yielded more to the segregationist arguments expressed by the accomplice. Contrary to expectation, when no roles were assigned, *S*s did not indicate less esteem for the accomplice than was the case when roles were assigned.

It is concluded that role assignments have an important influence on group members' perceptions of one another. Consequently, role assignments are important determinants of members' responses to one another's behavior.

REFERENCES

BERLO, D. Allocation of procedural responsibilities as a determinant of group productivity and satisfaction. Unpublished doctoral dissertation, Univer. Illinois, 1956.

BORGATTA, E. F., & BALES, R. F. Interaction of individuals in reconstituted groups. *Sociometry*, 1953, **16**, 302–320.

CARTER, L., HAYTHORN, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1950, **46**, 589–595.

JANIS, I. L., & KING, B. T. The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49**, 211–218.

OSGOOD, C. E., & SUCI, G. S. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.

SMITH, E. E. Effects of threat induced by ambiguous role expectations on defensiveness and productivity in small groups. *Group Process Lab. tech. Rep.*, 1956, No. 1. (Nonr-1147(03), NR 170 226, Univer. Colorado) (Mimeo)

STEINER, I. D., & DODGE, J. S. Interpersonal perception and role structure as determinants of group and individual efficiency. *Hum. Relat.*, 1956, **9**, 467–480.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 246-254

VOCABULARY KNOWLEDGE AND USAGE OF SCHIZOPHRENIC SUBJECTS:

A SIX-YEAR FOLLOW-UP[1]

LOUIS J. MORAN[2]
Veterans Administration Hospital, Houston, Texas

DONALD R. GORHAM[3]
Veterans Administration Hospital, Perry Point, Maryland

AND WAYNE H. HOLTZMAN
University of Texas

FOR THE past 30 years vocabulary performance, used as a measure of former intelligence level, has played a prominent role in the assessment of intellectual deterioration. In recent years, however, the use of vocabulary measures in this manner has provoked a barrage of criticisms (see Yates, 1956, for a comprehensive review). Foremost among these criticisms, as they apply to deterioration in schizophrenics, has been the contention that vocabulary performance also is impaired in schizophrenics, and, hence, cannot be used as a so-called "stable baseline" from which to judge deterioration in other abilities.

Impairment of vocabulary performance in schizophrenics, if present, could be manifested as a decline coincident with and a function of the psychotic episode itself, or as a gradual erosion with continued illness and institutionalization, or both. In an earlier investigation (Moran, 1953) the first possibility was examined in a cross-sectional study of 40 matched pairs of schizophrenic and normal subjects (*S*s). It was concluded that "... although the schizophrenic may appear to 'define' words as adequately as the control subject, he is less able than the control subject to attain the 'more difficult solutions' of further differentiating the word meanings, or to use the same words in reasoning or in communicating" (p. 16).

The present study reports the results of a follow-up of these schizophrenic *S*s. Yates (1956) has concluded, after a careful review of available literature, that "... a potent (perhaps the most important) factor in inducing a decline in vocabulary is the length of time a patient is institutionalized" (p. 436). The present study tests this hypothesis by a re-examination of the same schizophrenic *S*s with the same battery of word meaning measures after six additional years of institutionalization.

In addition, the current "social adequacy" of the *S*s was rated in the present study. Thus, it was possible also to evaluate the prognostic utility of word meaning measures by relating test performance in 1952 to rated social adequacy in 1958.

In 1952, 54 schizophrenic *S*s completed the entire test battery and one *S* completed all but the word association test. The samples described in the present report were drawn from this total of 55. During the ensuing 6 years, 35 patients remained hospitalized, 5 died, and 12 were discharged and living in the community. The other 3 *S*s could not be located.

USE AND UNDERSTANDING OF WORD MEANINGS IN 1952 AND 1958

Since the hypothesis to be tested here concerned the effect of prolonged institutionalization upon the schizophrenics' use and understanding of word meanings, the 35 *S*s still hospitalized after six years were of primary interest.[4] A full description of this sample may be found in the earlier report (Moran, 1953). In brief, in 1952, all had been diagnosed as chronic paranoid schizophrenic without known organic disease and with an age range of 24 to 57 years ($M = 41.3$), and had been institutionalized from 1 to 19 years ($M = 4.1$). Of the 35 hospitalized patients, 30 were retested with the same battery of word meaning measures used in 1952.[5]

---

[1] This study was made possible by a grant from The Hogg Foundation for Mental Health, University of Texas. Grateful appreciation is expressed to Ruth M. Hubbard and other members of the Waco VAH staff for their cooperation in carrying out this study.

[2] On leave from University of Texas.

[3] Formerly at Waco VAH.

[4] It would, of course, be of interest to know whether performance of the discharged group had changed in the six years. Several of these *S*s refused to be retested, however, and others were in remote parts of the coun-

Word Meaning Measures

The test battery consisted of the Wechsler-Bellevue (WB) vocabulary subtest, Form 1, and an additional group of seven special word meaning measures. The seven special tests (untimed) used a common core of 25 very familiar words and were designed to afford measurements, for each word, of the following dimensions of word usage and understanding.

1. *Definition.* The first 10 words (Part I), which were names of things, were scored separately from the last 15 words (Part II), which had no "thing" referent. Each of the 25 definitions was scored for (*a*) adequacy and (*b*) abstractness.
2. *Synonym* (*Recall*). *S* gave as many synonyms as he could think of for each word, and performance was scored for percentage correct synonyms given.
3. *Synonym* (*Recognition*). The task was to underline the appropriate synonyms among eight alternates for each word, and was scored for percentage correct synonyms underlined.
4. *Sentence Construction.* The 25 words were grouped into 12 sets, ranging from one to three words each. The task was to write a meaningful single sentence for each set, using all of the words in the set, and was scored for total correct sentences written.
5. *Similarities.* Other combinations of the 25 words were used in a 17-item similarities task. Part I consisted of 7 items using only words with a "thing" referent; the 10 items in Part II used only words without a "thing" referent. Each of the 17 items was scored for (*a*) adequacy and (*b*) abstractness.
6. *Analogies.* An 11-item analogies test, constructed around the basic 25 words, was scored for total correct.
7. *Word Association. S* gave his first association (one word) to each of the 25 words. Associations were scored on a 4-point scale for relatedness, i.e., "conceptual distance" of response word from the stimulus word.

The purpose of this special battery of tests was to permit an intensive multidimensional examination of the *S*'s knowledge and usage of a few very familiar words. Thus, for each of the 25 words, a measure was provided for (*a*) ability to define, (*b*) conceptual level of definition, (*c*) "breadth" of understanding (number of other words available to express the same meaning as that of the stimulus word), (*d*) preciseness of understanding (percentage correct synonyms), (*e*) ability to communicate (sentence construction), (*f*) ability to form concepts (similarities), (*g*) conceptual level of concepts (similarities), (*h*) ability to reason symbolically (analogies), and (*i*) associative links.

The test forms and detailed scoring instructions for these measures are available from ADI.[6] Every effort was made to administer the tests in the same manner as in 1952. All protocols (the original 30 of the retest 30) were then shuffled and scored independently by two scorers, using the scoring manual constructed for the original study. Interscorer reliabilities on the different measures ranged from .94 to .98. After scoring reliabilities were computed, all scoring differences were resolved jointly by the two scorers in close accord with the manual.

TEST-RETEST RESULTS

No significant ($p < .05$) mean changes occurred in the six-year period, although variability was somewhat greater on second testing, especially on the adequate definitions task. On the whole, test-retest reliabilities were high, considering the small number of items on most tests and the years intervening between performances. The overall picture is one of remarkable stability in the use and understanding of word meanings.

It has been argued that the scoring system for the WB vocabulary subtest may be insensitive to subtle qualitative changes in definitions which are alleged to be charac-

---

try. Three *S*s living in the community were retested. Inspection of their test-retest scores revealed very slight changes, with inconsistent direction on different tests in the battery.

[5] Five hospitalized *S*s were not retested solely because they were in remote regions of the country. Thus, their omission suggests no a priori reasons for bias in the retest sample.

[6] A copy of the test forms described in this study, with scoring manual, has been deposited with the ADI. Order Document No. 4355 from the ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $3.75 for microfilm or $2.00 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

COMPARISON OF PERFORMANCE BY HOSPITALIZED SCHIZOPHRENICS ON WORD MEANING MEASURES IN 1952 AND 1958

(No significant mean differences; $n = 30$)[a]

Word Meaning Measure	1952		1958		r[b]
	M	*SD*	*M*	*SD*	
Wechsler-Bellevue Vocabulary subtest, Form 1	21.8	6.86	20.4	7.22	.93*
Other Word Meaning Measures					
Definitions					
Adequate definitions, total	24.2	1.02	22.4	3.93	.25
Abstract definitions, total	13.0	5.49	13.0	5.91	.80*
Adequate definitions (Part I)	9.9	.35	9.7	.88	
Adequate definitions (Part II)	14.3	.94	12.7	3.15	
Abstract definitions (Part I)	6.0	3.05	6.3	2.85	
Abstract definitions (Part II)	7.5	3.33	6.7	3.64	
Synonym (Recall)					
Percentage correct synonyms	51.5	18.01	51.1	18.37	.75*
Synonym (Recognition)					
Percentage correct synonyms	73.5	14.43	72.9	15.37	.76*
Sentence construction					
Adequate sentences	7.9	2.55	8.5	2.64	.31
Similarities					
Adequate, total	9.9	3.78	9.1	4.75	.77*
Abstract, total	7.0	4.23	6.9	4.92	.79*
Adequate (Part I)	5.1	1.39	4.7	1.76	
Adequate (Part II)	4.6	2.58	4.4	3.15	
Abstract (Part I)	3.2	1.94	3.2	2.15	
Abstract (Part II)	3.6	2.66	3.6	2.76	
Percentage adequate scored abstract	65.7	20.19	66.7	25.38	
Analogies					
Number correct	6.4	2.91	6.2	2.83	.73*
Word Association[c]					
Relatedness score	58.8	9.69	57.2	10.26	.48*

[a] Mean differences computed by direct difference t for correlated means. Adopted significance level $p < .05$.

[b] Pearson product-moment correlation between first and second testing.

[c] $n = 29$, one *S* did not complete task.

* $p < .05$; correlation significantly different from zero.

TABLE 2

QUALITATIVE SCORES ON WECHSLER-BELLEVUE VOCABULARY SUBTEST, FORM 1, IN 1952 AND 1958

(No significant mean differences, $n = 30$)[a]

Type of Definition	1952		1958		r[b]
	M	*SD*	*M*	*SD*	
Synonym	11.9	5.53	12.3	5.26	.91*
Use and Description	3.0	1.23	2.8	1.15	.58*
Explanation	2.4	2.02	3.3	2.38	.64*
Inferior	2.6	1.75	1.6	1.36	.68*

[a] Mean differences computed by direct difference t for correlated means. Adopted significance level $p < .05$.

[b] Pearson product-moment correlation between first and second testing.

* $p < .05$.

teristic of schizophrenics. Feifel (1949) applied a 5-point scoring system for quality of definition to the Terman Vocabulary List and demonstrated significant mean differences in quality scores between "abnormal" (organic, manic-depressive, schizophrenic) and normal *S*s matched for Terman vocabulary scored in the standard manner. However, essentially the same qualitative scale, used with the WB vocabulary subtest, Form 1, failed to differentiate schizophrenics from normal *S*s in a matched-pair design with *S*s equated for WB vocabulary scored in the standard manner (Moran, Moran, & Blake, 1952).

To determine whether shifts in qualitative level of definition on the WB vocabulary had occurred in the present sample, all definitions, test and retest, were scored independently by two judges in terms of four qualitative levels: (*a*) synonym, (*b*) use or description, (*c*) explanation, and (*d*) inferior. The full rationale of this scoring system, with specific criteria and examples for each category, may be found in Moran (Moran, et al., 1952). Interscorer agreement on qualitative categories was 96%.

The stability of WB vocabulary performance in this sample of schizophrenics, even when scored for quality of definitions, is evident in Table 2. No significant mean differences occurred at any qualitative level. Again, the high retest reliability of the major category, synonyms, attested to the consistency of vocabulary performance over the six-year period. As a further check, qualitative categories were assigned weights of 3 for synonyms, 2 for use or description, 1 for explanation, and 0 for inferior. The weighted scores on test-retest yielded a direct difference t of .23, $p > .70$.

However, still another possibility must be considered. Although no mean differences on the word meaning measures were statistically significant, it may be observed in Table 1 that 13 of the 19 sample means were lower on second testing, though usually by an extremely small margin. A remote possibility still remained that statistically significant differences in direction of change were obscured in comparisons of mean performance levels. To test this hypothesis, each of the eight word meaning measures was rescored for direction of change. On each test, *S* received a 0 if his second score was lower than the first, 1 if the same, and 2 if higher. Six of the measures evidenced no consistent change in direction. However, this weighting system exposed a significant downward shift in WB vocabulary and in the 25-word definitions test.

On the WB vocabulary test, 19 *S*s made a lower score on second testing, 3 the same score, and 8 a higher score. The difference between the observed distribution and that under the hypothesis of no directional change was significant ($p < .05$) as tested by chi square corrected for discontinuity. On the 25-word definitions test (total adequate), 17 *S*s made a lower score on second testing, 9 the same score, and 4 a higher score. This downward shift in scores was also statistically significant.

To summarize at this point: (*a*) 19 comparisons by means of direct difference *t*'s between raw scores in 1952 and 1958 failed to demonstrate a single significant change in mean level of performance on any measure; (*b*) a comparison of 1952 and 1958 performance on WB vocabulary, scored for quality of definition, showed no significant mean differences; but (*c*) when magnitude of change was ignored and direction of change emphasized, a statistically significant decline in performance was found in WB vocabulary and in the 25-word definitions test.

Obviously, the change in WB vocabulary performance was an extremely subtle one. A clue was sought in the study of "off-quadrant" cases, i.e., those below the median on one testing and above on the other. However, as might be evident from the high test-retest correlation ($r = .93$), this attempt was thwarted by the paucity of cases in the off quadrants (one *S* in each).

Qualitatively, one of the most notable differences between the first and second testing was an increased verbosity. The mean number of words used to define the words in the WB vocabulary test increased almost twofold in 1958 (from 119 to 197). An illustration of this tendency is provided in Table 3, which lists definitions to the word *join* by every *S* in 1952 and 1958.

TABLE 3

DEFINITIONS OF THE WORD *Join* IN 1952 AND 1958

Subject	1952	1958
1	Join a church, army	That's when you shake hands with someone and you become in unity
2	Join together	To join something, some kind of lodge or something
3	Put together	Put together or get together
4	Sign up with some kind of organization	Connect up with something else
5	Tie	Connect
6	Cooperation of people, join an organization or hands to be married	Like join a church, a crowd joined together
7	Join something	I don't know
8	To put together, assemble	To put together
9	Together	Put together
10	Anything put together	Join organization, put anything together
11	Put together	Get with; place together or make connections
12	Together	Put together
13	Join clubs	Means to join together
14	To enter into an organization	To become a member of
15	Baptize; connect	Connect
16	To enlist	To enlist; put two things together
17	Join something together	Like join anything together
18	Together	Connect
19	Unite	Bring together
20	Unite	Like putting two pieces together
21	To unite	Join the church
22	Unite	Put together; don't know the correct definition
23	Connect	To put together
24	Join a church	Like join something together
25	Attach	Draw close together; get together
26	Connected	Going together
27	Join a club	Like join the Red Cross; join together; a group of people joined together
28	Unite	To become affiliated with
29	Put together	To make an alliance with; put together; bring together
30	Together	To put together or to join a church or secret organization

It might be hypothesized that the subtle decline in vocabulary performance noted in these *S*s was a function primarily of lessened facility in communicating their knowledge rather than of loss of knowledge per se. This hypothesis is supported by the increased difficulty these *S*s had in expressing themselves, as evidenced by the increase in number of words used to define test words, and, qualitatively, by the strained expressions illustrated in Table 3. The words in the 25-word definitions test were selected deliberately because of their common everyday usage. Failure to define some of these words adequately, while defining more difficult WB vocabulary words, also supports the hypothesis.[7] Further, the only decline in performance noted in the entire test battery was specific to the two definitions tests, which placed a premium upon verbal facility.

If the very subtle changes on the WB vocabulary test represented a loss of knowledge or failure of memory, more credit would be lost on words at the upper end of the scale. Recall of the meaning of rarely used words like *moiety* or *traduce*, for example, would be more tenuous than for words like *cedar* or *tint*. On the other hand, if the decline represented a failure to communicate knowledge, credit might more easily be lost at any point along the scale. Thus, a test of the hypothesis that the decline was due to difficulty in communicating knowledge rather than loss of knowledge was afforded by examination of the distribution of the 1952–1958 gains and losses in credit for Words 1 through 42 on the WB vocabulary test.[8]

The frequency of loss or gain in credit points (-1, $-\frac{1}{2}$, 0, $+\frac{1}{2}$, $+1$) was tallied for each word in the WB vocabulary test. The distribution of gains and losses throughout the 42 words was highly similar for each word. On Words 1–21, credit was lost at second testing 91 times and gained 51 times; on Words 22–42, credit was lost 82 times and gained 55 times. The hypothesis of loss in verbal facility as opposed to loss of "knowledge" thus received confirmation.

Whatever the nature of this slight drop in performance on definitions tests, what effect would it have upon a "deterioration index" of the type that employed the discrepancy in performance between vocabulary tests and problem solving tasks? For illustrative purposes, a Deterioration Index (DI) of this type was constructed by dividing WB vocabulary raw score into the summed raw scores on analogies and adequate similarities. The resulting ratio thus served as an index of the discrepancy between "former ability level" (size of vocabulary) and present "mental efficiency" (problem solving ability).

The mean DI for the normal *S*s who served as a matched-pair control group in 1952 (see Moran, 1953) was .899. The mean DI in 1952 for the present sample of schizophrenics was .748. This mean difference of .151 was significant ($p < .001$). Thus, such an index in 1952 would have shown the schizophrenic group to be deteriorated.

An index constructed in the same manner from 1958 scores of the schizophrenic group yielded a mean of .750, almost identical to the .748 of 1952. The correlation between the 1952 DI and 1958 DI was .23, $p > .05$. Thus, although individuals changed in relative position, the group as a whole showed no evidence of deterioration by this operation in the six years.

However, in this instance actual former ability level (at least in 1952) was known, making it unnecessary to use inferences from present vocabulary performance. Therefore, a DI was constructed using the actually observed former level (1952 vocabulary performance) in a ratio with current "mental efficiency" (1958 analogies and similarities). The mean of this DI, .686, was significantly ($p < .001$) lower than the DI of .750 obtained when 1958 vocabulary level was contrasted with 1958 performance on analogies and similarities. Thus, actually observed "deterioration" was undetected by a deterioration index of the traditional type.[9]

---

[7] It should be noted, however, that on this test the only possible direction of change was downward. The mean on first testing was 24.2, where the maximum possible score was 25. Hence, the regression here may be more statistical than psychological.

[8] Scoring criteria on the WB vocabulary test are progressively more stringent through the list. Thus, greater precision in the choice of words is required of definitions for words at the upper end of the scale. Also, fewer *S*s reach the upper end of the scale, so that an instance of credit loss carries proportionately more weight than an instance at the lower end of the scale. Both of these factors, however, are working against the successful outcome of the hypothesis as tested here.

[9] Correlations of each of the three DIs with age and number of years institutionalized were not statistically significant.

DISCUSSION AND CONCLUSIONS

Viewed collectively, the results of this study demonstrated a remarkable stability of performance by hospitalized chronic schizophrenics on word meaning measures taken six years apart. Performance on the kinds of tasks commonly thought to be sensitive to deterioration, e.g., analogies and similarities, showed no significant change in this period. Paradoxically, the only change noted was in the one measure thought to be relatively insensitive to deterioration—definitions. But even here the decline was so small as to go undetected by some of the usual statistical tests.

The downward shift in WB vocabulary score on second testing, though very slight, occurred in about two thirds of the *S*s. The uniformity of the change was apparent also from the test-retest correlation of .93, i.e., rank order was maintained closely on first and second testing. Decreased verbal facility, rather than loss of "knowledge," seemed to be a primary reason for the poorer performance on second testing.

Was this slight loss in verbal facility due to increased schizophrenic pathology, to the stultifying effect of six additional years of institutionalization, to the normal process of aging, or to a complex interaction of these and possibly other variables? Unfortunately all of these possibilities are inextricably tied together in the present sample. However, certain lines of evidence may be pursued which might tend to implicate some of these variables more than others.

First, if the decline in verbal facility were due solely to increased schizophrenic pathology, it should be reflected most strongly in those tests most "diagnostic" of schizophrenic thinking. Yet observed mean scores on tests that in 1952 reliably differentiated this same schizophrenic group from matched normal *S*s changed proportionately less than WB vocabulary observed mean score (ignoring for the moment the fact that these observed mean changes were not statistically significant). Furthermore, a significant number of *S*s showed a slight drop in WB score, but not on the more "diagnostic" tests. It does not seem plausible that performance on measures known to be sensitive to schizophrenic thinking should be unaffected by increased schizophrenic pathology, whereas performance on a measure known to be highly insensitive in this respect showed a significant decline.

Second, if institutionalization per se led to decreased verbal facility, performance level and years of institutionalization should tend to be correlated. The correlation between years hospitalized and WB vocabulary score in 1952 was .01. A more sensitive test would be the relationship between years institutionalized and amount of *change* in WB vocabulary score in the six year period. This correlation was .03. Since in 1952 this sample had been hospitalized from 1 to 19 years, in 1958 the range was from 6 to 25 years. Of those now hospitalized from 9 to 25 years, nine had a lower score on second testing and six a higher score. Of those now hospitalized from 6 to 8 years, however, nine had a lower retest score and only three a higher score. This difference, though not significant, is contrary to Yates' (1956) hypothesis that length of institutionalization is the most important factor in inducing a decline in vocabulary performance.

Adequate assessment of the role of aging in the observed slight decline would require comparable data on normal *S*s which, unfortunately, are unavailable. Since *all S*s were now six years older, a measure of the effect of aging by relating age to amount of change would test for rate of change at different ages, not amount of change due to aging. Thus, although rate of change did not differ significantly with different ages in the present sample, consistent change in the entire group as a function of normal aging seems, by process of elimination, to be the most plausible explanation of the observed slight decline in verbal facility.

However subtle the change in performance on the word meaning measures was in the six years, and whatever the explanation for it, it is clear that changes even as slight as those observed in the present sample can vitiate the use of current vocabulary performance as an estimate of former ability level for use in a deterioration index. This was illustrated by the comparison of deterioration indices constructed separately from performance in 1952 and 1958, which indicated no change, although a similar index contrasting actual former ability level (1952 vocabulary) with present mental efficiency (1958 analogies and similarities) revealed a significant "deterioration" in the six-year period.

The conclusions drawn from re-examination of the use and understanding of word meanings by these chronic schizophrenics, after six additional years of hospitalization, were as follows:

1. Overall performance on word meaning measures was remarkably stable.
2. Performance on definitions tests tended to drop very slightly but consistently among the *S*s.
3. This subtle decline in performance on definitions tests seemed more likely due to slightly decreased verbal facility or communication of knowledge, rather than to loss of knowledge per se.
4. Loss in verbal facility seemed more plausibly attributed (by process of elimination) to normal aging processes than to increased schizophrenic pathology or to length of time institutionalized.
5. A deterioration index, using the discrepancy between current vocabulary test performance and current performance on problem solving tasks, completely obscured an observed significant deterioration in performance.
6. Impairment of vocabulary performance in schizophrenics seems to be manifest as a decline coincident with and a function of the psychotic episode itself, rather than as a gradual erosion with continued illness and institutionalization.

Prognostic Utility of Word Meaning Measures

Word meaning measures of the type described in the previous section, though frequently used in various ways for psychodiagnosis, rarely have been tested empirically for prognostic value. In the present study it was possible to locate 47 of the 50 surviving *S*s. Thus, an opportunity was afforded to compare directly the current adjustment of the *S*s with their performance on the word meaning measures six years earlier.

Two criteria of current adjustment were available. First, 12 *S*s had been discharged and were living in the community, as against 35 who were still hospitalized. Although discharge from an NP hospital may be conditioned by a great many sociological factors having little to do with the patient's psychiatric condition, it was presumed that discharge could be used as a moderately reliable criterion of significant improvement. Second, the social adequacy of each *S* in his current environment was evaluated by an experienced social worker. The Pinchak-Rollins Social Adequacy Rating Scale (SARS) was used for this purpose.

Social Adequacy Ratings[10]

The SARS provides a five-point scale for each of eight areas of adjustment. Separate forms are provided for rating adjustment in the hospital and in the community, with slight changes in wording to make the rating appropriate to the setting. Detailed instructions are given, with a summary of approximately 100 words describing the nature and scope of the variable to be rated. Each of the five points on a variable carries a sentence describing the range and types of behavior to be rated for that particular area of adjustment. Variables also are scored 1, 2, or 3 for confidence in the rating.

The eight areas of rated adequacy are given below. The wording in parentheses was substituted for that in italics when the scale was used to rate hospitalized *S*s.

A. Responsibility for money (or canteen coupons). The extent to which the individual is realistically responsible about his money (or coupons), its source, and its use.

B. Personal appearance. The extent to which the individual has been able to maintain his personal appearance, e.g., dressing, shaving, oral hygiene, haircuts, grooming, etc., in conformity with *those of others in his environment* (expectations for other persons in society).

C. Personal habits. The extent to which the individual has been able to maintain *standards of his immediate environment* as to eating, sleeping, and bathing (social standards).

---

[10] We are in the debt of Louis E. Pinchak and Glen W. Rollins, Social Work Service of the Waco Veterans Administration Hospital, for use of the SARS and for the great contribution of Pinchak in locating *S*s throughout the United States and personally supervising the SARS ratings. A comprehensive report on the SARS is in press (Pinchak & Rollins, 1959). The separate scales for adjustment in the community and in the hospital, as used in the present study, have been deposited with the ADI. Order Document No. 6318 from the ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

D. Responsibility for vocational (educational or industrial assignments). The extent to which the individual has been able to maintain continuing responsibility for productive work (on his assignment).

E. Social group attendance. The extent to which the individual assumes responsibility for attending (ward or hospital) social, recreational, religious, and avocational activities.

F. Social group participation. The extent to which the individual is responsible for emotional interaction in a social group (on the ward or in special hospital groups).

G. Responsibility for family and immediate associates (other patients and personnel). The extent to which the individual shows responsibility for the rights and well-being of *members of family and immediate associates in his environment at home, at church, on the job, etc.*, (other patients, ward personnel, and other associates).

H. Interpersonal relations. The extent to which the individual maintains sustained relationships with others (patients and personnel) on a person-to-person basis, demonstrating personal-emotional involvement.

The SARC was used by social workers in the Veterans Administration to rate the 35 patients in the various hospitals in which they resided. The 12 former patients were rated in their local community by social workers in the Veterans Administration and in collaborating agencies. The majority of *S*s were in Texas, but a sizable minority were scattered throughout the United States. For 12 judges independently rating the same 10 *S*s, Pinchak and Rollins (1959) report an average intercorrelation of .89 on SARS total score.

Prediction of Social Adequacy in 1958

The extremely poor prognosis for almost all of the *S*s in the present sample was one of the most immediate and striking findings in this follow-up. Only 12 of the original sample had been discharged within the six year period, and of these only one was judged to be making an adequate adjustment and one a borderline adjustment. The other 10 discharged *S*s were found to require active supervision by family and associates in order to remain in the community.

Since virtually none of the *S*s had recovered, the prognostic efficiency of the word meaning measures was tested against an extremely restricted criterion, the degree of illness within a pathological group rather than pathology vs. recovery. Despite this restriction, it may be seen in Table 4 that performance in 1952 on all of the word meaning measures showed a slight positive correlation with social adequacy in 1958, whether judged by discharge or by the SARS ratings. Although virtually none of the individual correlations were statistically significant, the fact that all 27 correlations were in the predicted positive direction strongly suggested the existence of a very slight but definite relationship.

TABLE 4

PREDICTION OF SOCIAL ADEQUACY IN 1958 FROM PERFORMANCE ON WORD MEANING MEASURES IN 1952

Word Meaning Measure	Correlation with 1952 Test Performance		
	Discharge[a] from Hospital ($n = 47$)	Hospital Adjustment[b] ($n = 35$)	Outside Adjustment[b] ($n = 12$)
Wechsler-Bellevue Vocabulary subtest, Form 1	.31	.16	.30
Other Word Meaning Measures:			
Abstract definition, total	.35	.40*	.11
Synonym (recall) percentage correct	.26	.28	.23
Synonym (recognition) percentage correct	.27	.17	.29
Sentence construction, total correct	.17	.09	.40
Adequate similarities, total	.37	.16	.19
Abstract similarities, total	.38	.12	.35
Analogies, total correct	.32	.22	.34
Word association, relatedness[c]	.32	.19	.19

[a] Biserial r, $\sigma r_b = .198$.

[b] Rank order correlation coefficients.

[c] $n = 46$, 34, and 12, respectively; one hospitalized *S* did not complete test.

* $p < .05$.

The deterioration index constructed from performance in 1952 did not correlate significantly with discharge from the hospital or with rated social adequacy. Age and years institutionalized correlated −.03 and .04, respectively, with rated social adequacy ($n = 47$).

It may be concluded that performance on the word meaning measures in 1952 showed a very slight but apparently genuine relationship to discharge from the hospital and to rated social adequacy in 1958, but that a deterioration index, using a ratio of the same measures, was unrelated to either criterion.

SUMMARY

A sample of 55 chronic paranoid schizophrenics that had been examined in 1952 with an extended battery of word meaning measures was followed up in 1958. Thirty of the 35 *S*s still hospitalized were readministered the same test battery. Overall performance was remarkably stable over the six year period, with no significant mean change on any measure. However, performance on definitions showed a slight decline in most *S*s. This decline seemed due to loss of verbal facility rather than to loss of knowledge, or memory failure. Also, the decline seemed associated more plausibly with normal aging than with increased schizophrenic pathology or length of time institutionalized. A deterioration index, using a ratio of WB vocabulary to analogies plus similarities was demonstrated to obscure an observed "deterioration" in performance.

The 35 hospitalized *S*s and 12 discharged *S*s also were rated for social adequacy in 1958. Only two of the discharged *S*s were judged to be making an adequate adjustment. Test performance in 1952 showed an extremely low but positive relationship to discharge from the hospital and to rated social adequacy. A deterioration index, constructed from the same tests, showed no relation to either criterion.

REFERENCES

FEIFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genet. psychol. Monogr.*, 1949, **39**, 151–204.

MORAN, L. J., MORAN, F. A., & BLAKE, R. R. An investigation of the vocabulary performance of schizophrenics: II. Conceptual level of definitions. *J. genet. Psychol.*, 1952, **80**, 107–132.

MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, **67**(20, Whole No. 370).

PINCHAK, L. E., & ROLLINS, G. W. The social adequacy rating scale, a preliminary report. In press.

YATES, A. J. The use of vocabulary in the measurement of intellectual deterioration: A review. *J. ment. Sci.*, 1956, **102**, 409–440.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 255-262

A MICROGENETIC APPROACH TO THE AMNESIC SYNDROME[1]

GEORGE A. TALLAND
Harvard Medical School

In Korsakoff's psychosis the memory defect is so outstanding, so much more severe than other symptoms that the disease is customarily referred to as the amnesic syndrome. Experimental studies (Talland, 1958b, 1959a) have thrown more light on the cognitive impairment characteristic of this syndrome, suggesting dysfunctions in basic processes other than that which is broadly known as memory. Korsakoff patients function defectively in at least three major areas of cognitive processes: (*a*) in the selective operations which determine perceptual discrimination and the categorization of concepts, (*b*) in reorienting their initial attitude towards a task or situation confronting them, and (*c*) in their capacity to integrate sequential experiences into unitary structures. Quite possibly, these three phenotypically distinct processes are functionally related to one another and also to the processes by which experiences become available for voluntary recall. The tendency to isolate fragmentary events and observations would seem to be particularly relevant to the twofold amnesic disturbance of Korsakoff patients, i.e., their virtual inability to learn or recall information received and events witnessed after the onset of the disease, and their diminished capacity to remember those which had happened and which they had known before. Impairment in temporal integration was a defect characteristic of the two extreme cases of amnesia reported in the literature (Conrad, 1953; Störring, 1931), and has been experimentally demonstrated with a group of less severely damaged Korsakoff patients (Talland, 1958a). The relevance of this defect to the amnesic symptoms was suggested also by Lidz (1942) and is at least implicit in Bürger-Prinz and Kaila's (1930) graphic definition of the Korsakoff patient as consisting of cross sections without a longitudinal continuity.

In order to be available for recall, experiences must become embedded in some cognitive systems, and quite obviously this happens as a neurophysiological process in the brain. Common experience, as well as studies of such damaged functioning as occurs in the Korsakoff syndrome, shows that clear perception is not a sufficient condition for accurate recall or for voluntary recall of even approximate exactitude. Remembering depends on an additional process of embedding which is developmentally subsequent to perception. In the intact brain, the complete process of perception is likely to include the phase of fixation, but in abnormal conditions these two processes may become divorced, and perception itself may not reach full development. An incompletely structured percept is also inadequately structured within the cognitive system which subserves recall.

Perception is usually thought of as an instantaneous event, for the process rarely admits analysis into stages. The Gestalt school was particularly emphatic on the immediacy of the experience, although Bartlett (1916), Michotte (1927), and Sander (1927) demonstrated that objective analysis need not imply a return to Wundtian introspection and argued in favor of studying perception as a developmental process. Hebb's (1949) associationist theory also treats perception as a progressive process and is particularly relevant to the case of Korsakoff's syndrome, in that its neurological model offers an explanation why an incompletely structured percept should be lost to memory. Most of the experimental work and theoretical formulations directed to the developmental study of cognitive processes have been done in Sander's wake, and are now best known by Werner's (1956) translation of *Aktualgenese* as the microgenetic approach (Flavell & Draguns, 1957). The basic tenet of this approach is that perception, like concept formation or problem solving, is a process of development from the diffuse to the structured, progressing over a microscopic temporal sequence. Normally, the successive phases of this development are indistinguishable; only the end product can be inspected. However, under suitable experimental conditions the progress of microgenesis can be brought under observation.

---

[1] This research was supported in part by a Research Grant, M-767(c) from the National Institute of Mental Health, United States Public Health Service.

G. Talland is connected with the Psychology Laboratory, Department of Psychiatry, Massachusetts General Hospital, supported in part by Air Force Contract AF-49-(638)-98; the Merrill Trust; the Harrington Trust, Harvard Medical School.

The purpose of the present study was to test certain hypotheses which follow from the foregoing formulation of the cognitive disorder in Korsakoff's syndrome, by investigating the microgenetic process of perception. Several findings concerning the rigid adherence of the Korsakoff patients to a particular orientation or set (Bürger-Prinz & Kaila, 1930; Grünthal, 1924; Müller-Suur, 1949; Talland & Miller, 1959) point to an impairment of their microgenetic processes, for these must be corrective as well as cumulative. The incapacity to integrate sequences of input may also extend to the order of intervals over which percepts develop microgenetically, intervals so short that memory effects are excluded from their range by definition. More particularly, the argument for an impairment of the microgenetic process of perception in the Korsakoff syndrome would be supported if the following hypotheses were to be confirmed:

1. In Korsakoff patients, under conditions of progressively increased exposures, the perception of ambiguous figures progresses towards an accurate and structured percept at a slower rate and less consistently than in controls.
2. With Korsakoff patients, successive exposure of the same ambiguous figure does not result in a more consistent or more accurate approach to perceptual structuring than does presentation repeated after intervals.

Structuring, in the context of these two hypotheses, means the integration and elaboration of parts or details into whole figures; consistent stands for the gradual development of this process, accurate for its veridicality.

METHOD

Design. The study consisted of three experiments. The purpose of the first two was to test the two hypotheses stated above, by providing the conditions for a gradual structuring of percepts: one over a more extended period of time with intervening activity, the other in a continuous process. Experiment III, in which no control group participated, attempted to sort out the effects of repeated exposure and length of exposure on the microgenetic progress of perception.

Subjects. The experimental sample consisted of 16 patients, 4 men and 12 women, diagnosed by neurologists as having Korsakoff's psychosis. They were in the chronic phase of the disease, having been at the Boston State Hospital for periods ranging from 3 to 15 years or longer. They all had a history of alcoholism prior to hospitalization which, in the first place, was for the neurological symptoms of Wernicke's encephalopathy.

There were two control groups, one consisting of 15 and the other of 10 psychiatric and neurological patients whose sociocultural background was similar to that of the Korsakoff patients, but who were otherwise not matched with them.

Material. The five composite figures shown in Fig. 1 were used in all the experiments. The first four are legendary beasts: (A) the winged bull of ancient Assyria and of the Evangelist St. Luke; (B) the hippogriff, a composite of horse, eagle, and lion; (C) merman and mermaid; (D) the Etruscan hippocamp, half horse and half dolphin. They were copied with some modifications from a book on fabulous monsters (Lum 1951). Fig. 1E is a caricature[2] of the Chevalier d'Eon, French diplomat and famed duelist, whose transvestite habits gained him considerable notoriety in London society of the 1770s (White, 1952). These inherently ambiguous designs were chosen for their relative incongruity and unfamiliarity, thus creating conditions for a more extended process of recognition by demanding the repeated checking of observations and the reconciliation of incoherent details. Photographs of the drawings, none of which was over 7 in. in its longer, nor less than 4½ in. in the shorter dimension, were mounted on cardboard and viewed by *S* binocularly in a Gerbrands tachistoscope at a distance of about 24 in. from his eyes, in the center of a field of vision of 7¾ × 7¾ in.

Procedure. In all the experiments, *S* was instructed to be alert, watch, and report what the picture was he had seen flashing in the instrument. He was told that the time for inspection would be very brief, but received no clue in advance of the type of figure he would be shown, nor subsequently any suggestion of what it may have represented. However, if he reported nothing or if his response was very meager, *E* prompted him with such nondirective questions as "Is that all?", "Anything else?" or by repeating his words with an inflection of incompleteness.

In Experiment I the composite figures were shown under the conditions of Repeated Presentation. The five figures were presented in standard sequence from A to E, first with an exposure of .1 sec., then again in the same order with exposures of .25, .5, and 1.0 sec. All 16 Korsakoff patients and 15 control *Ss* participated in this procedure. Experiment II was conducted under the conditions of Successive Presentation, in which each figure was shown four times with exposures of .1, .25, .5, and 1.0 sec., before presenting the next one. Fourteen Korsakoff patients and a control group of 10 *Ss*, different from those in the first control group, participated. The reason for

---

[2] The composite figure of the Chevalier d'Eon originally appeared in the September 1777 issue of the *London Magazine*.

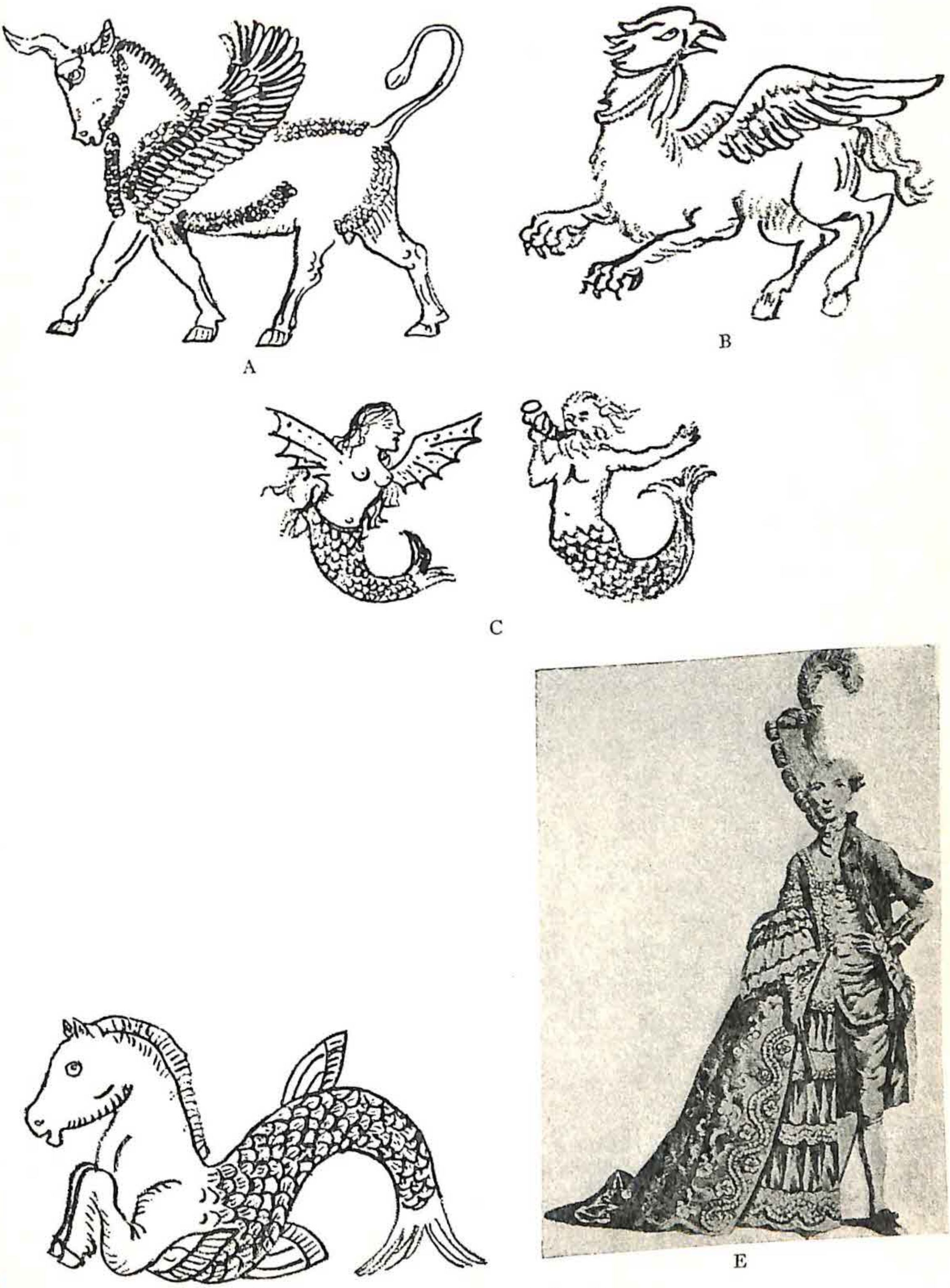

FIG. 1. Composite figures.

A. Assyrian winged bull
B. Hippogriff
C. Merman and mermaid
D. Etruscan hippocamp
E. Chevalier d'Eon

introducing the two experimental conditions in that order was to protect against a spurious support of the second hypothesis. Since the two experiments were conducted six months apart, it was unlikely that any memory effect would carry over from the first to the second occasion in the amnesic patients. Nevertheless, if such an effect should prevail, it would not favor perceptual structuring under conditions of Repeated Presentation. In the control sample, of course, the operation of such an effect was quite probable, and this was the reason for selecting different *S*s for the two tests.

Experiment III was administered to the 14 Korsakoff patients four weeks later, using again the method of Successive Presentation. This time, each of the seven figures was shown with a different schedule; one seven times with exposures of .01, .01, .01, .1, .25, .5, and 1.0 sec.; another six times with exposures of .1, .1, .1, .25, .5, and 1.0 sec.; a third five times with exposures of .25, .25, .25, .5, and 1.0 sec.; a fourth with exposures of .5, .5, .5, and 1.0 sec.; the last three times, each for a full second. The schedule of presentation was shifted across the five figures so that in the group of 14 *S*s each design was shown three times with four schedules and twice with a fifth. Last, the figures were presented one by one outside the tachistoscope, so that *S* could inspect them for as long as he cared to before describing each. Full inspection was allowed to the control group at the end of Experiment II.

Scoring. The data used in the analysis of results were not directly derived from the experimental procedure. Responses to the figures were recorded verbatim, and subsequently tabulated under several headings. In principle, this procedure is subject to the risk of systematic error; however, the main coding categories used in this instance represented three so clearly defined universes that, after the establishment of certain guiding rules, hardly any occasion for error could occur. For responses to the first presentation of a figure the protocols were subdivided into correct, partly correct, or incorrect. Horse for Fig. 1A was incorrect, for Fig. 1B partly correct, as was lion, eagle, or bird, but not rooster. Responses to subsequent presentations were distinguished according to their being Elaborations, Changes, or Replications of a previous description, and then divided into correct and incorrect. Elaboration included all modifications of a previous response in the direction of a structured percept, a new detail, or the correction of a detail, as well as a more specific definition instead of a general one (e. g., horse for animal, two fishes for fish). Change meant the replacement of a specific response by another which could be correct (e.g., mermaids for butterflies) or incorrect (e.g., Napoleon for George Washington). Replication could either be a verbatim repetition of a previous response, or words amounting to much the same (e.g., circus horse for trick horse), or simply the phrase "same" beyond which *S* would not expand. These three classes do not exhaust all the responses; the residual includes all instances of impoverishment, i.e., a response sparser in content yet no more specific than a previous description of the same figure, meaningless answers, comments like "black and white," and finally failure to report anything at all.

RESULTS

In the tabulation of the results only four headings were used because the instances of Change were too few to allow for further splitting; and the significance of Same is not contingent on the accuracy of the previous response, for no matter how correct that may have been it invariably allowed for considerable further Elaboration. Since the analysis is concerned with trends in the direction of perceptual structuring, the tables do not include scores based on responses to the first presentation of a figure. For the purposes of tabulation, responses to the subsequent presentations of the five composite figures are combined into single scores, and these are expressed as a percentage of the highest possible total for any one *S*, i.e., 15 in Experiments I and II.

With an exposure of .1 sec. one Korsakoff patient failed to see anything articulate in four composite figures on both first and second testing, another failed likewise with three figures on the first occasion and did not take part in the experiment with Successive presentation. Apart from these instances, defective vision or the shortness of the exposure did not interfere with the perception of some distinct figure in either experimental or control groups. Most members of the latter gave a correct or partly correct answer after the briefest exposure of Fig. 1B, C, D, and E, but only three recognized Fig. 1A as a bull or a nonspecific animal in the first and second experiment alike. The same number of Korsakoff patients responded more or less correctly to the first exposure of Fig. 1A in both Experiments I and II, as compared with half of the group on the first presentation of Fig. 1C, and about two-thirds with Fig. 1B, D, and E. Almost all of them gave a correct or partly correct description of the last four figures on subsequent presentations.

The tendency in the two groups to Elaborate, Change, or Replicate previous descriptions can be traced from the data in Table 1, and then compared for the conditions of presentation. Differences between the mean scores of the Korsakoff and control groups, between the performances of the two control groups, and of the Korsakoff patients under the two conditions of presentation have been tested for

TABLE 1

MEAN ELABORATION, CORRECT ELABORATION, CHANGE, AND REPLICATION SCORES OF THE KORSAKOFF AND CONTROL GROUPS IN EXPERIMENTS I AND II AFTER SECOND, THIRD, AND FOURTH PRESENTATIONS OF THE COMBINED FIVE COMPOSITE FIGURES

(Scores express the actual response as a percentage of the possible total)

	Elaboration			Correct Elaboration			Change			Replication		
	Kors.	Con.	*t* Ratio	Kors.	Con.	*t* Ratio	Kors.	Con.	*t* Ratio	Kors.	Con.	*t* Ratio
Repeated Presentation	41	62	2.83**	23	41	3.37**	23	11	1.16	24	18	1.10
Successive Presentation	24	59	$z = 3.48$**	17	42	$z = 2.56$**	6	13	$z = 1.60$*	65	20	$z = 3.62$**
t ratio	4.85**	.26		2.35*	.34		4.11**	.35		11.50**	.28	

* $p < .05$.
** $p < .01$.

significance, using *t* ratio for correlated means in the last instance, and Mosteller and Bush's (1954) version of the Mann-Whitney test when variances are not homogeneous. This test yields a standard score (z).

It is apparent that in both experimental conditions the rate of total Elaboration and of correct Elaboration is lower in the Korsakoff than in the control group, but that the former Change their description significantly less often and Replicate it more frequently only under conditions of Successive Presentation. The two control groups differ in none of these aspects of their performance, though one was presented with the composite figures *repeatedly* and the other *successively*. The Korsakoff group's total and correct Elaboration and Change were significantly higher, its Replication was significantly lower under conditions of repeated presentation.

The results of Experiment III are presented in Table 2. Neither Elaboration nor Replication seems to be a direct function of the number of successive presentations, and neither varies with the mean length of exposure in a direction which would suggest that—at any rate, within the range of 1 sec.—the briefness of exposure time accounted for the low rate at which Korsakoff patients evolved their percepts. Table 2 does not show but the records furnish further evidence in support of this negative conclusion. In the longer sequences, Elaboration occurred as often after a short as it did after a long exposure, and many of the Replications came towards the end of the series, i.e., following a longer exposure. Also, in accordance with this conclusion is the absence of significant differences between any of the mean scores in Experiment III and the corresponding scores in Experiment II.

TABLE 2

MEAN ELABORATION, CORRECT ELABORATION, CHANGE, AND REPLICATION SCORES OF THE KORSAKOFF GROUP IN EXPERIMENT III, FOLLOWING THE SECOND AND SUBSEQUENT PRESENTATIONS OF THE FIVE COMPOSITE FIGURES

(Scores express the actual response as a percentage of the possible total)

Number of successive exposures	3	4	5	6	7	All
Elaboration	18	31	21	33	31	28
Correct Elaboration	11	21	14	26	23	20
Change	7	7	5	9	4	6
Replication	68	55	64	53	56	58

Inspection outside the tachistoscope elicited from the Korsakoff patients barely more descriptive material, as judged by number of words or units of report, than tachistoscopic presentation; nevertheless, 80% of it involved some Elaboration and 70% was veridical, all the errors occurring with Fig. 1A. In the control group, 78% of the responses to prolonged inspection constituted some Elaboration of previous descriptions, all correct.

As a matter of interest it may be noted that, excepting correct Elaboration of Fig. 1A, the contribution of the five figures to the four sets of scores was remarkably even in all experiments and with both groups of *S*s. Figures 1B and C provided the best material for Elaboration, Fig. 1D and E for Replication, Fig. 1A for incorrect Elaboration or Change. Figure 1E also inspired many Changes, more particularly from one sex to the other, but also across the periods of history. In repeated presentations, one *S* recognized it first as Napoleon and then as George Washington, another in a different group as a gladiator and Don Juan. Other specific identifications included Queen Elizabeth, Abraham Lincoln, Martha Washington, Scots dancer with bagpipes, Chinaman, Jew with locks of hair, Spanish dancer, and knight in armor. Seven control and three Korsakoff *S*s recognized the combination of man and woman in this figure after Successive tachistoscopic presentation, and so did two additional members of either group following longer inspection. The remaining *S*s assimilated the incongruous details of the image perceived, or avoided the identification of its sex. When asked whether it was a man or woman, their answers sounded definite enough, though in some instances supported by reasoning of this type: "it must be a woman because of the feathers."

DISCUSSION

The experiments confirm the hypotheses tested. The rate at which Korsakoff patients structure ambiguous figures into percepts is slower, and under Successive Presentation it is also more often interrupted by phases of stagnation, than in the control group. Experimental conditions designed to display the microgenetic development of perception actually show less progress in Korsakoff patients than do conditions which interrupt this process; there is no difference between the effects of the two conditions on the control groups. The Korsakoff patients' failure in the progressive structuring of percepts and in the gradual correct identification of the composite figures may therefore not be attributed to a primary memory defect, for this would not be manifested in the brief intervals between two Successive Presentations of a figure, and would certainly be more effective over the longer period which elapses between Repeated Presentations, quite apart from the interference the intervening tasks might or might not produce (Talland, 1959b).

The differential results of the experiments with composite figures seem paradoxical only if the amnesic defect is regarded as the primary psychological symptom of Korsakoff's psychosis. They agree with other observations which demonstrate the inability of those patients to reorient themselves to a situation, to modify their initial perception of or attitude to it. Likewise, on account of their rigid adherence to the first percept, successive presentations, even well above the threshold of clear discrimination, make for very slight increments in structuring or, as the results of correct Elaboration show, in approximating a veridical percept. In contrast, Repeated Presentation by interrupting the first set adopted towards a composite figure, advances the process of structuring and correction at a faster and steadier rate. Of course, the interruption of a set can only indirectly advance the process of perception. In the absence of any transfer of information from one exposure of a figure to its Repeated Presentation, the *S* would only be in the same position he initially held.

It must therefore be inferred that some memory effects did operate in Korsakoff patients, making for a gradual advance in their perception both towards completeness and accuracy. The alternative explanation would attribute this progress to the increased exposure time with each Repeated Presentation. Experiment III, though not decisive on this issue, would argue in favor of the first explanation. It shows that the length of exposure time—within the range deemed suitable for exploring the microgenesis of perception—is not directly related to the rate of Elaboration or the likelihood of Replication. The responses given by the Korsakoff patients after prolonged inspection of the figures were significant not so much for their brevity or rapid closure, as for containing some elaboration in spite of the paucity of content, in almost every instance. The quality of their reports following tachistoscopic presentation could therefore not be attributed to any impediment in verbal fluency.

A genuine deficit in the microgenetic development of perception thus seems fairly well established in the amnesic syndrome. This observation would obviously gain in significance if it were theoretically related to other, clinically more apparent, symptoms of the disease. Kragh (1955), discussing the microgenesis of perception, suggested that possibly every sequence of development must be repeated a number of times before it can acquire sufficient strength to result in full-fledged perception. He illustrated the completion of a perceptual process with a negative instance drawn from Schumann's (1907) experiment in which, after brief exposure, *S*s were known to report either that they could not recognize the word even though they had seen the letters clearly, or that they had recognized it but could not recall it. The microgenetic paradigm of perception fits neatly into Hebb's (1949) model of simple cell assemblies aggregating into phase sequences over a period of time. In this model, learning proceeds and memories become available the same way as perception occurs, i.e., by means of the establishment of facilitations in the nervous system, though at different levels of organization. The effect of these facilitatory channels is to increase the probability that certain events in the brain follow other events, that a percept attain completion, that a concept be formed, that a memory recur when sought. If it is assumed that in Korsakoff patients these facilitations fail to become established, the probabilistic model can account for those instances in which their recall, recognition, or other cognitive processes function efficiently, as well as for the many more occasions in which they do not. The model also accommodates the various observations concerning the Korsakoff patients' relative incapacity to shift expectancies, and certainly agrees with the paradigm of isolated experiences proposed above. It has the virtue of comprehensiveness, but is inherently no more convincing than an inhibitory model. Several observations of the Korsakoff patients' behavior, more particularly of the kind of verbal performance which is often referred to as confabulation, fits more closely a model of defective inhibitory mechanisms. There is no necessity to choose either model at the expense of the other, for it is reasonable to believe that a disease which interferes with the efficient functioning of facilitatory systems also damages the inhibitory processes. Neither is the microgenetic approach to cognition tied to any one neurological model; its purpose in this study was to demonstrate the applicability to the psychopathology of Korsakoff's syndrome of such purely psychological constructs as are the isolation of single experiences and the relative fixity of originally adopted attitudes.

SUMMARY

The microgenetic process of perception was examined in a group of Korsakoff patients by means of serial tachistoscopic presentation of ambiguous drawing with successively longer exposure. The drawings consisted of unfamiliar composite figures.

Under the experimental condition of successive presentation, which is the most favorable for observing this process, the gradual structuring and elaboration of percepts proceeded at a slower rate than it did when other tasks intervened between the two exposures of the same figure. Under both conditions the process of structuring and elaboration was slower and less complete with the Korsakoff patients than in control groups of psychiatric and neurological patients. The number of repeated exposures and length of inspection did not determine the rate of structuring and elaboration.

The differential effect of the Korsakoff patients' performance under the two conditions of presentation supports the view that the perceptual defect in Korsakoff's syndrome is not secondary to a memory disorder, and that an isolation of rudimentary impressions and an impaired capacity to revise attitudes are general psychological symptoms of this disease.

The findings based on this microgenetic approach and of other cognitive studies of Korsakoff's syndrome were considered in the light of Hebb's neurophysiological model.

REFERENCES

BARTLETT, F. C. An experimental study of some problems of perceiving and imaging. *Brit. J. Psychol.*, 1916, **8**, 222–266.

BÜRGER-PRINZ, H., & KAILA, M. Über die Stuktur des amnestischen Symptomenkomplexes. *Z. Neurol. Psychiat.*, 1930, **124**, 553–595.

Conrad, K. Über einen Fall von "Minuten Gedaächtins." *Arch. Psychiat.*, 1953, **190**, 471–502.

Flavell, J. H., & Draguns, J. A microgenetic approach to perception and thought. *Psychol. Bull.*, 1957, **54**, 197–217.

Grünthal, E. Über das Symptom der Einstellungsstörung bei exogenen Psychosen. *Z. Neurol. Psychiat.*, 1924, **92**, 255–266.

Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.

Kragh, U. *The actual-genetic model of perception-personality*. Lund: Gleerup, 1955.

Lidz, T. The amnestic syndrome. *Arch. Neurol. Psychiat.*, 1942, **47**, 588–605.

Lum, P. *Fabulous beasts*. New York: Pantheon, 1951.

Michotte, A. Rapport sur la perception des formes. *Proc. 8 Int. Congr. Psychol.*, 166–174. Groningen: Noordhoff, 1927.

Mosteller, R., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. I. Cambridge, Mass: Addison-Wesley, 1954. Pp. 289–334.

Müller-Suur, H. Beitrag zur Frage des Korsakov-Syndroms und der Analyse der amnestisch-strukturellen Demenz. *Arch. Psychiat. Nerv.*, 1949, **181**, 683–711.

Sander, F. Über Gestaltqualitäten. *Proc. 8 Int Congr. Psychol.*, 183–189. Groningen: Noordhoff, 1927.

Schumann, F. Psychologie des Lesens. *Bericht über den II Kongress für experimentelle Psychologie in Würzburg*. Leipzig: Barth, 1907. Pp. 153–183.

Störring, G. Über den ersten reinen Fall eines Menschen mit völligem isoliertem Verlust der Merkfähigkeit. *Arch. Psychol.*, 1931, **81**, 257–284.

Talland, G. A. Perception of apparent movement in in Korsakoff's psychosis. *J. Pers.*, 1958, **26**, 337–348. (a)

Talland, G. A. Psychological studies of Korsakoff's psychosis: II. Perceptual functions. *J. nerv. ment. Dis.*, 1958, **127**, 197–219. (b)

Talland, G. A. Psychological studies of Korsakoff's psychosis: III. Concept formation. *J. nerv. ment. Dis.*, 1959, **128**, 214–225. (a)

Talland, G. A. The interference theory of forgetting and the amnesic syndrome. *J. abnorm. soc. Psychol.*, 1959, **59**, 10–16. (b)

Talland, G. A., & Miller, Alicemarie. Perceptual sets in Korsakoff's psychosis. *J. abnorm. soc. Psychol.*, 1959, **58**, 234–240.

Werner, H. Microgenesis and aphasia. *J. abnorm. soc. Psychol.*, 1956, **52**, 347–353.

White, T. H. *The scandalmonger*. New York: Putnam, 1952.

(Received August 6, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 263-270

THE ADAPTIVE FUNCTIONS OF WIT AND HUMOR[1]

WALTER E. O'CONNELL[2]
University of Texas

THE PURPOSE of this study[3] was to test psychoanalytic formulations pertaining to wit and humor as personality variables related to the adjustment of the individual.

While the words "wit" and "humor" have been used almost interchangeably by theorists with a nonpsychoanalytic orientation, workers with a Freudian background have tended to perceive differences between the two concepts. These distinctions appear in the feelings expressed indirectly through the jest and in the degree of personal adjustment reflected by preferences among jests.

Appreciation of humor, in the restrictive meaning offered by Freud, is associated with empathy that still is not overwhelmed by the misfortunes of others, with suitably flexible emotionality, with little use of repression, and with tolerance for oneself and others (Flugel, 1947; Freud, 1905, 1928; Grotjahn, 1957; Kramer, 1958). The form of jest labelled humor presents an objectively stressful situation in which effective flight and fight are both relatively impossible. The chief character in the Freudian humor situation then reacts to this inescapable stress with neither resignation nor overt hostility but with a humorous retort.

Wit has been regarded as a means of indirect expression for latent hostile urges. Freud was never certain whether this mechanism was essentially pathological or adaptive. Early in his thinking he correlated the appreciation of hostile wit with psychoneurotic symptoms (1905), but in later years (1921) he included it among normal adaptive mechanisms.

Harmless or nonsense wit employs the same verbal mechanisms (e.g., play on words, displacement of accent) as humor and hostile wit, but the situational settings of the other two forms are missing. Therefore in nonsense wit there is neither jesting response to adversity nor disguised expression of hostility.

Little research has been directed toward testing the Freudian concepts of wit and humor. This dearth of studies is anomalous in the light of the increased attention being given to humor as a requisite for the psychotherapist, the diagnostician, and, generally, the mature person. Occasionally, humor has been mentioned by psychoanalytically-oriented investigators but its connotations have been confused with those of wit, which has been somewhat more frequently studied (Redlich, Levine, & Sohler, 1950). Researchers have generally postulated an inverse relationship between repression and overt behavior but a direct relationship between repression and appreciation of wit. Thus, the person who "represses" hostility should not act in an overtly hostile manner, but should enjoy hostile wit.

There are a number of theoretical and experimental obstacles to study in this area. Operational measures of repression and concomitant behaviors fail to take into account the complex and dynamic theoretical background of these concepts. Theoretically a person could repress an impulse yet show such impulse derivatives as hostile behavior or appreciation of wit in his overt actions. An adequate plan for testing the relationship between repression and appreciation of wit requires manipulation of such variables as the strength and extent of repressions, the psychological "distance" of the derivatives from the repressed impulse, and the tension generated both by external stressors and by intrapsychic conflicts (Freud, 1915). Within this theoretical framework, wit could be considered as a tension-reducing mechanism, with appreciation of wit contingent upon "some" but not "too much" tension or stress. The present study was planned as an initial test of such complex interrelationships.

---

[1] Based upon a doctoral dissertation submitted to the University of Texas, 1958. The author is grateful to Ira Iscoe for his guidance in this research.

Presented at APA meetings in Cincinnati, Ohio, September 1959.

[2] Now at Veterans Administration Hospital, Waco, Texas.

[3] 92-page copies of complete study, including the test items, have been deposited with University Microfilms, Inc., 313 North First Street, Ann Arbor, Michigan. Order Document No. Mic 58-3993, remitting \$2.00 for 35 mm. microfilm or \$3.68 for 5½ by 8½ in. photocopies.

Hypothesis I. In a situation where hostile impulses are aroused and direct means for expressing them are blocked, hostile wit might provide an indirect outlet for the hostile drives, thereby reducing to some degree both the underlying urge and the mental energy required to block entry of the impulse into consciousness and/or overt behavior. Since a sudden reduction of energy or tension is experienced as pleasurable according to psychoanalytic theory, the vehicle for the reduction (the joke) should also be appreciated. Therefore, the first prediction is that groups of subjects exposed to stress appreciate wit more than comparable groups tested under nonstressful conditions.

Hypothesis II. Psychoanalytic theory has been strongly oriented toward the psychology of the male. Hence, theoretical deductions as to the role of wit and humor in female development are difficult to find. Grotjahn (1957) stated that in this culture hostile wit is the wit of the male rather than of the female. If the show of hostility even in derivative forms is not allowed to the female, one might predict that nonsense or "harmless wit" would find more favor with women than with men.

Hypothesis III. From a review of the development of the Freudian concept of wit one can make certain predictions about preferences for hostile wit, depending upon the interaction of the variables of adjustment (internal stress) and presence or absence of objective stress. The maladjusted person, presumably with strong unacceptable hostile impulses, employs the facade of the hostile joke for pleasurable and partial impulse gratification. But when the external situation involves a hostile atmosphere the repressions of the maladjusted person are exaggerated to such an extent that disguised release is not possible. On the other hand, the more adjusted individual, supposedly not burdened with excessive repressions (or intrapsychic tensions) can enjoy hostile wit mechanisms for tension reduction when objective stresses increase. Thus, it is predicted that under nonstressful conditions the maladjusted person should appreciate hostile wit more than the well adjusted person, but when stressors are applied, the well adjusted subject should appreciate hostile wit more than the maladjusted person.

Hypothesis IV. In contrast to his attitude toward wit, Freud always extolled the humorist as disdaining the use of excessive repressions, and he depicted humor as the most adaptive nonpathological defense. Other writers have concurred with Freud's picture of the humorist as being mature, nonhostile, and loving toward himself and others. One would predict, then, more frequent appreciation of humor in well adjusted persons.

Hypothesis V. In comparison with humor and hostile wit, less has been written on the subject of nonsense or harmless wit. This form of the comic, resting more upon the techniques of wit than upon savings in countercathexis, was given less emphasis in Freud's theoretical system (1905). There is some consensus that a person who can allow himself relaxation and release through the appreciation of nonsense wit is not overwhelmed by his problems (Kris, 1952; Levine & Redlich, 1955). Therefore, the final hypothesis states that a well adjusted person appreciates nonsense wit more than a poorly adjusted one.

METHOD

Selection of Independent Test Variable

In order to test Hypotheses III, IV, and V, a measure was needed that would separate the *S*s into groups according to their level of adjustment. The Self-Activity Inventory (SAI), which gives an absolute discrepancy score between the *S*s' self-rating and his ideal rating on 54 items relating to psychological needs, was selected as the independent test variable (Worchel, 1957). The rationale for the use of the absolute discrepancy score between self and ideal (S-I score) is provided by self-theory (Rogers, 1954), according to which tension is experienced when a person is aware of the discrepancy between his self-concept and his ideal concept. Therefore, it was predicted that the greater the discrepancy (high S-I score) the greater the anxiety or tension experienced. A number of studies have indicated that the S-I score is inversely related to the level of adjustment and of stress tolerance, and significantly correlated with psychometric tests of anxiety (Worchel, 1957). The reliability of this scale by a test-retest method over an eight-week interval was .79 for the self-concept and .72 for the ideal concept.

Selection of Dependent Test Variable

The rationale for choice of a dependent variable arises from Freud's belief that passive appreciation of wit and humor echoes the psychic states of the jest maker. Thus, a test for the appreciation of wit and humor should provide an indicator of the defensive use of these devices. Two existing tests have attracted considerable interest as research tools in the field of the comic but neither follow the Freudian paradigm of humor (Cattell & Luborsky, 1947; Redlich, Levine, & Sohler, 1950). For the present study a Wit and Humor Appreciation Test was developed around forms of jests that follow deductions from psychoanalytic theory.

Test construction required the suitable definition of wit and humor, selection of a pool of items that would reasonably be subsumed under these categories, and decision as to the type of response to be recorded.

A pool of 65 jokes were assembled and typed separately on filing cards. Eleven judges (8 male and 3 female clinical psychologists) were asked to classify the 65 items in private as Humor, Hostile Wit, or Nonsense Wit. To avoid idiosyncratic definitions of these concepts, each judge was provided with a resume of Freud's ideas concerning wit and humor. Of the original pool of 65 items, 32 were unanimously classified as belonging in one of the three categories. Since 10 were in the humor category and 11 were categorized both as Hostile Wit and as Nonsense Wit, one item was eliminated randomly from each of the last two classes. At this point, there was an assortment of 30 jokes (10 each of Humor, Nonsense Wit, and Hostile Wit) available for inclusion in the Wit and Humor Appreciation Test (WHAT). Examples of items are as follows:

> In various times punishment has been cruel and frequently ineffective in arousing fear. There were men who endured the greatest tortures without bowing to its influence. One man who was to be hanged was carried to the gallows in a basket by four guards. He was carried because his arms and legs had been cut off. In this situation he called laughingly to the crowd present: "I never thought that a poor devil like me would be carried about by four servants." (Humor)
>
> While Emperor Augustus was traveling through his domains, he noticed a man in the crowds who bore a striking resemblance to himself. He beckoned to him to come over and asked:
>
> "Was your mother ever employed in my home?"
>
> "No, sire," replied the man, "but my father was." (Hostile Wit)
>
> Description of a cow followed by a couple of ducks: Milk and Quackers. (Nonsense Wit)

*S*s were requested to record their preference for each joke on a scale from 0 ("dislike very much") to 4 ("like very much"). The rationale for such a scale of preference comes from psychoanalytic studies of wit and humor that regard "implicit laughter" and appreciation as the key response, rather than more overt measures.

Several pilot studies were conducted to measure the effect of extraneous variables upon test responses. To minimize testing difficulties random presentation was arranged, i.e., six jokes on each of five sheets, the sheets in random order for each *S*; on each sheet two items from each wit and humor category, counterbalanced for position on the page.

Subjects

The 332 *S*s were college students, essentially homogeneous in regard to age and class, enrolled in seven introductory courses in psychology, sociology, or anthropology.

Procedure

Two test scores were needed for each *S* in order to evaluate hypotheses relevant to adjustment and appreciation of wit and humor. To test the predictions concerning appreciation of wit and humor under stress, certain classes were designated as stress groups, and others as nonstress groups. For the stress groups, two classes with a total of 150 *S*s were selected, both sections in introductory anthropology meeting 2 hours apart. The nonstress groups contained 182 *S*s from five intact classes. The testing of both groups was within the regular classroom hour in two consecutive class sessions.

For both stress and nonstress groups the *E* presented the SAI, supposedly as part of the routine testing procedure of the University, and returned during the next class session with the Wit and Humor Appreciation Test.

For the two stress classes, a faculty member accompanied the *E* on the second day and became the instigator of both failure-stress and insult techniques. He introduced himself as a professor of clinical psychology who was shocked and appalled at the attitude of the class in taking the SAI the previous class period. The results of the test, he said, pointed out that the members were either lying about themselves or were grossly maladjusted. He also stated that since the tests reflected a marked lack of adjustment, the majority of the class should give serious consideration either to leaving college or to changing their majors. After reiterating the insults and threats a few times, the instigator informed the students that they would be given one more test, and that they should display some maturity in taking it, rather than performing like high school freshmen. They were also admonished against communicating with their neighbors in any way. The *E* distributed the WHAT immediately upon the conclusion of the instigator's remarks.

After the tests were completed and collected, the aim of the experiment was explained to the *S*s. To aid in the release of whatever tensions might have accumulated, about 15 min. were spent encouraging the *S*s to give their views and feelings regarding the stress experience.

In order to provide the *E* with objective evidence of class behavior to supplement his own observations, eight undergraduate students were asked to observe and note in writing the reactions of the classes. All eight observers were unaware of the purpose of the experiment but concurred unanimously in their views as to the widespread feelings of anxiety and suppressed hostility in these stressed classes.

RESULTS

Hypothesis I stated that the groups of persons exposed to stress would appreciate wit

TABLE 1

MEAN APPRECIATION RESPONSES FOR WIT AND HUMOR ITEMS

Sex	Humor			Hostile Wit		Nonsense Wit		Total	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Male	196	17.7	5.81	24.6	6.12	17.2	6.60	59.5	15.92
Female	136	17.9	4.68	22.7	6.81	20.2	6.75	60.8	16.43
t		0.36		2.63*		4.11**		1.18	

* $p < .01$, one-tailed test.

** $p < .001$, one-tailed test.

more than comparable groups tested under nonstress conditions. The mean scores for the stress and nonstress groups on Hostile Wit as well as Nonsense Wit and Humor were quite similar under both conditions. Hypothesis I, therefore, was not confirmed.

Hypothesis II concerned another group difference, that of a sex difference in preference for Hostile as compared with Nonsense Wit. Statistical analysis confirmed the hypothesis that males appreciate Hostile Wit more than do females ($t = 2.63$, $p < .01$, one-tailed test), while women favor Nonsense Wit ($t = 4.11$, $p < .001$).

The remaining hypotheses were formulated with reference to the S-I scores of adjustment. In order to compare the mean variables of S-I scores, sex, and the presence or absence of stress, a 2 × 2 × 2 analysis of variance design was selected. To assign *S*s to the required eight groups, all S-I scores were plotted in ascending order. From this total of 332 scores, 111 *S*s were grouped in both high and low S-I groups, to test the extremes of adjustment in a normal group. The low group included S-I scores from 0 to 39 ($M = 25.6$, $s = 8.12$), while the high S-I groups had scores from 57 to 122 ($M = 75.1$, $s = 11.96$). From both S-I groups, stressed and nonstressed *S*s were separated, then these four groups were split into male and female groups. The smallest number of *S*s for any of the eight groups was 17 in the high S-I, nonstressed, female group; hence this number became the standard number of *S*s for each group. To arrive at an *N* of 17 in the seven remaining groups, *S*s with the highest S-I scores were eliminated from the four low S-I groups, while those with the lowest S-I discrepancies were removed from the four high S-I groups.

Hypothesis III stated that the maladjusted *S*s (as measured by high S-I scores) would appreciate Hostile Wit more than those of good adjustment (low S-I) under nonstress, but that the relationship would be reversed under stressful conditions. For the dependent variable appreciation of Hostile Wit (Table 2), the simple interaction of S-I × Stressor ($F = 6.23$, $p < .05$) indicated that the low S-I group under stress and the high S-I group under nonstress appreciated Hostile Wit more than the combination of low S-I in nonstress and high S-I in stress situations. Hypothesis III can be considered to be confirmed by this interaction. There is also another simple interaction that was significant, S-I × Sex ($F = 9.51$, $p < .01$). This interaction showed that women with low S-I scores and men with high S-I scores preferred Hostile Wit more than the opposite combination of high S-I females and low S-I males. The triple interaction of S-I × Stress × Sex was likewise significant at the .01 level ($F = 9.33$). When tested under nonstress or "resting" conditions the high S-I males and the low S-I females both showed stronger preference for Hostile Wit than the contrasting groups. The two extremes under this situation are the very high hostile preference for the high S-I man and the very low appreciation for Hostile Wit evinced by the high S-I woman. With the introduction of stress, the low S-I men appreciated Hostile Wit more than the high S-I men. Among the women, it was the low S-I group that appreciated Hostile Wit under this condition.

TABLE 2

ANALYSIS OF VARIANCE OF PREFERENCE FOR HOSTILE WIT

Source of Variation	*df*	Mean Square	*F*
S-I Discrepancy	1	31.0	1.01
Stressor Conditions	1	3.1	
Sex	1	45.8	1.50
S-I × Stressor	1	190.5	6.23*
S-I × Sex	1	291.0	9.51**
Stressor × Sex	1	48.2	1.58
S-I × Stressor × Sex	1	285.4	9.33**
Error	128	30.6	
Total	135		

* Significant at the .05 level.

** Significant at the .01 level.

TABLE 3

ANALYSIS OF VARIANCE OF PREFERENCE FOR HUMOR

Source of Variation	*df*	Mean Square	*F*
S-I Discrepancy	1	176.7	7.21**
Stressor Conditions	1	69.2	2.82
Sex	1	0.3	
S-I × Stressor	1	6.0	
S-I × Sex	1	45.8	1.90
Stressor × Sex	1	1.5	
S-I × Stressor × Sex	1	43.5	1.81
Error	128	24.6	
Total	135		

** Significant at the .01 level.

TABLE 4

ANALYSIS OF VARIANCE OF PREFERENCE FOR NONSENSE WIT

Source of Variation	*df*	Mean Square	*F*
S-I Discrepancy	1	47.1	1.20
Stressor Conditions	1	2.3	
Sex	1	423.5	10.83**
S-I × Stressor	1	1.2	
S-I × Sex	1	95.6	2.44
Stressor × Sex	1	67.9	1.74
S-I × Stressor × Sex	1	49.4	1.26
Error	128	39.1	
Total	135		

** Significant at .01 level.

Hypothesis IV stated that well adjusted individuals have a greater appreciation for humor than do maladjusted people. One *F* ratio, that of S-I scores, was large enough to be significant. That ratio was in line with the prediction of the hypothesis, and reflects the differential in preference for Humor of the low and high S-I groups (Means of 18.8 and 16.5). Thus, Hypothesis IV can be considered confirmed.

The appreciation of Nonsense Wit was the topic of Hypothesis V, the assumption being that those high in adjustment would appreciate Nonsense Wit more than the less adjusted individuals. The resulting *F* ratio did not approach significance at the .05 level. The sex difference was the only significant *F* ratio, with means of 17.0 and 20.6 for men and women, respectively.

To test the reliability of each component of the dependent variable two classes were randomly selected, 43 males and 27 females. By the split-half method reliabilities (corrected, Spearman-Brown) for the component variables were: Nonsense Wit, .84; Humor, .77; and Hostile Wit, .61. While not extremely high, the reliabilities are acceptable when it is considered that the population was fairly homogeneous and that there were only five items in each half of the test.

DISCUSSION

The results of this study show that appreciation of wit and humor is amenable to reliable measurement, and that groups separated on the basis of sex, stress, and adjustment have recourse to these devices under differing circumstances. It, therefore, seems likely that attempts to explain wit and humor preferences by one factor alone give an incomplete picture of the dynamics of wit and humor.

The first two hypotheses were concerned with testing stress and sex differences in the appreciation of wit and humor. As for Hypothesis I, no significant differences were found in wit and humor preferences between groups under stress and nonstress conditions. Hypothesis II was confirmed since males were found to appreciate Hostile Wit significantly more than did females. On the other hand, women liked Nonsense Wit far more than did men. This finding does not differ greatly from the results of the study of a college population by Landis and Ross (1933). In that study, men appreciated "repression" jokes, while the women preferred "nonsense" jokes. Since the populations tested in both studies were quite similar, although separated by 25 years in time, the sex difference in appreciation of wit appears to be rather stable. To what degree one can generalize beyond the youth of the middle class to different ages and classes is a problem for future research. It may be the expected pattern for the college male to enjoy the release of Hostile Wit while the college female appreciates the relatively motiveless pleasure of nonsense wit.

When the personality variable of adjustment was added, more interesting interactions were noted. Thus, there is some experimental evidence that measures of adjustment may be important in predicting appreciation of wit and humor. From the interaction of the self-ideal discrepancy with the measurement of

Hostile Wit certain features of Hostile Wit become clearer. That is, Hostile Wit appears to be predominantly a male reaction, and in everyday situations it is the more maladjusted male who is disposed to employ it. Yet under stress conditions the maladjusted males did not utilize this defense as much as did the well-adjusted males. These results are congruent with psychoanalytic theory, according to which the maladjusted person represses hostility, but can make use of the verbal distortions of Hostile Wit for partial release so long as the objective situation is not too stressful (Freud, 1915).

When the external situation has hostile implications, the maladjusted person seems to intensify his repressive efforts to the degree that wit is not so ready a release. On the other hand, the adjusted person can make greater use of Hostile Wit under situations of stress, since the additional weight of insult and failure threats is not so disabling to him. The findings regarding Hostile Wit also dovetail with the views of Alper (1957), who implied that the maladjusted individual uses defenses in ordinary situations to such an extreme extent that under stress he cannot call upon more intensive defense and is, therefore, more prone to disorganization. On the other hand, the well adjusted person can employ such defensive reserve under stress, since his stability is not strained during "resting" states.

Findings with hostile wit make possible a better interpretation of previous studies of wit. Certain researchers were ready to discount Freudian theory because of a positive relation noted between appreciation of Wit and hostile behaviors (Byrne, 1956; Luborsky & Cattell, 1947). In these studies, the key variables of adjustment and stress were not measured, and it was erroneously implied that psychoanalytic theory considers that the neurotic is not hostile (e.g., inverse relationship between repression and overt behavior). In this study it was found that maladjusted people, as defined by their self-ideal discrepancy scores, enjoy hostile wit under nonstressful situations. The findings of such investigators as Barry (1928) and Murray (1934) are in agreement. Similarly, Byrne (1956) discovered that among maladjusted hospitalized male patients seemingly not under stress, openly and subtly hostile subjects liked hostile cartoons.

Two *E*s seem to have allowed stress to enter their studies, although the designs did not originally call for it. Doris and Fierman (1955) found a significant difference in preference for hostile cartoons between *S*s who showed high and those who had low anxiety, but *only* when an opposite-sexed tester was present. In such a situation, the highly anxious *S* showed significantly less appreciation for hostile cartoons. It seems possible that a highly anxious person might be sufficiently stressed by the presence of a person of the opposite sex as a tester to lose his zest for hostile cartoons. Frankel (1953) believed she learned the problem areas of *S*s with the Blacky Test; she then administered jokes tapping such areas of maladjustment. She found that those *S*s judged to be maladjusted did not appreciate jokes reflecting the discovered problem area. Here again is the addition of stressor, this time specific to the area of maladjustment, accompanied by the lowering of appreciation for hostile wit. Frankel's method differed from that of the present study in that there was no attempt in this study to find jokes that were directed at individual problem areas.

Females seem to be an exception to the predictions of present theories. In the present experiment, it was the well adjusted girls who were able to take advantage of Hostile Wit under both stress and nonstress conditions. Perhaps if our culture frowns upon hostile wit for women, only certain types of women rely upon this mechanism for tension reduction.

It was found that, as predicted, well adjusted *S*s appreciated humor more than did the poorly adjusted. Since the responses to humor were not significantly related to sex or the presence of a stressor, and since the response to wit was more affected by the interaction factors, one might make the tentative assumption that appreciation of humor is a more stable personality characteristic than appreciation of wit.

One of the major weaknesses in this experiment pertains to the use of the S-I score as a measure of adjustment. At this point in the development of phenomenal measures, the evidence for the S-I discrepancy as a measure of the quality of a person's adjustment is not impressive. Further research is required before the question of the validity of the S-I as an indicator of adjustment is satisfactorily answered. Perhaps for the present a more guarded label for the S-I would be one of "phenomenal self-esteem." According to the Y phenomenon (Rogers, 1954) a small discrepancy between self and ideal scores could signify either a fairly successful adjustment or a highly defensive maladjustment. In the *E*'s opinion there is some doubt about the cogency of the psychological principle in this case. The low self-ideal discrepancies reported here (Mean 25.6) are not at a pathological extreme and average almost eight points higher than low self-ideal groups used elsewhere (Miller & Worchel, 1956). There is no doubt that a study of theoretical importance could be undertaken with *S*'s giving extremely congruent self-ideal ratings (e.g., discrepancy of less than 10 points). Such a group might be characterized by extreme repression, with very limited appreciation of humor (and possibly also of hostile wit).

An item analysis of the WHAT has raised further questions that make some of the initial findings inconclusive. Within the broad categories of wit and humor sex differences in preference have been discovered depending upon the amount and direction of hostility present in the jest and the object toward which the hostility is directed. These findings, to be reported later, have led the investigator to research in progress concerning the dimensions of Freudian humor.

Further studies of individual reactions under various types of stress are also required to discover whether wit and humor have abreactive value by themselves alone or only in combination with other tension-reducing behaviors.

SUMMARY

This study was undertaken to test the possible adaptive values of the Freudian mechanisms of Humor, Hostile Wit, and Nonsense (or "harmless") Wit, for *S*s grouped according to self-ideal discrepancies, sex difference, and the presence or absence of an external stressor.

Three hundred and thirty-two *S*s indicated their self- and ideal-adjustment on Worchel's Self-Activity Inventory. The *S*s then were presented with a Wit and Humor Appreciation Test, some under insult and failure (the stressor), some under nonstressful conditions.

Five hypotheses relevant to wit and humor were tested:

I. Groups exposed to the stressor appreciate wit and humor jokes more than comparable groups that are not tested under stress. This hypothesis was not confirmed since the stressed and nonstressed groups showed no significant differences in response to the test jokes.

II. Men appreciate Hostile Wit more than women do, while women prefer Nonsense Wit more than do men. This hypothesis was confirmed with resulting probabilities beyond the .01 point.

For the remaining three hypotheses a 2 × 2 × 2 analysis of variance was computed, with eight groups of 17 *S*s each, differentiated on the main variables of high and low self-ideal discrepancies, sex, and stressor—no stressor.

III. Maladjusted people appreciate hostile wit more than do well adjusted under nonstressful conditions. When a stessor is applied, well adjusted people appreciate Hostile Wit more than do the maladjusted. Hypothesis III was confirmed for the men only. The women showed a different pattern with the women of low self-ideal discrepancy (well adjusted) appreciating Hostile Wit more than the women of high self-ideal discrepancy (maladjusted), under both stressor and nonstressor conditions.

IV. The well adjusted person has a greater appreciation for humor than does the poorly adjusted person. The *F* ratio for the adjustment variable was significant at the .01 level, thus the hypothesis was confirmed.

V. The well adjusted person has a greater appreciation for Nonsense Wit than does the poorly adjusted person. This hypothesis was not confirmed.

It is concluded that the Freudian distinction between wit and humor is a valid one, with wit possibly operating to reduce tension under differing circumstances depending upon the interaction of adjustment, stress, and sex factors.

REFERENCES

ALPER, THELMA G. Predicting the direction of selective recall: Its relation to ego-strength and n achievement. *J. abnorm. soc. Psychol.*, **55**, 1957, 149–165.

BARRY, W. The role of subject-matter in individual differences in humor. *J. genet. Psychol.*, **35**, 1928, 112–128.

BYRNE, D. The relationship between humor and the expression of hostility. *J. abnorm. soc. Psychol.*, **53**, 1956, 84–89.

CATTELL, R., & LUBORSKY, L. Personality factors in response to humor. *J. abnorm. soc. Psychol.*, **42**, 1947, 402–421.

DORIS, J., & FIERMAN, ELLA. Humor and anxiety. *J. abnorm. soc. Psychol.*, **53**, 1955, 59–62.

FLUGEL, J. *Man, morals, and society.* New York: International Univer. Press, 1947.

FRANKEL, ESTER. An experimental study of psychoanalytic theories of humor. Unpublished doctoral dissertation, Ann Arbor: Univer. Michigan, 1953.

FREUD, S. Humor. *Int. J. Psychoanal.*, **9**, 1928, 1–6.

FREUD, S. Wit and its relation to the unconscious. (Originally published 1905) In A. Brill (Ed.), *The basic writings of Sigmund Freud.* New York: Modern Library, 1938. Pp. 694.

FREUD, S. Group psychology. (Originally published 1921) In J. Rickman & C. Brenner, *A general selection from the works of Sigmund Freud.* New York: Doubleday, Anchor Books, 1947. Pp. 169–209.

FREUD, S. Repression. (Originally published 1915) In *Collected papers.* Vol. III. London: Hogarth Press, 1950. Pp. 84–97.

GROTJAHN, M. *Beyond laughter.* New York: McGraw-Hill, 1957.

KRAMER, P. One of the preoedipal roots of the superego. *J. Amer. Psychoanal. Ass.*, **6**, 1958, 38–46.

KRIS, E. *Psychoanalytic explorations in art.* New York: International Univer. Press, 1952.

LANDIS, C., & ROSS, J. Humor and its relation to other personality tests. *J. soc. Psychol.*, **4**, 1933, 158–175.

LEVINE, J., & REDLICH, F. Failure to understand humor. *Psychoanal. Quart.*, **24**, 1955, 560–572.

LUBORSKY, L., & CATTELL, R. The validation of personality factors in humor. *J. Pers.*, **15**, 1947, 283–291.

MILLER, K., & WORCHEL, P. The effect of need-achievement and self-ideal discrepancy on performance. *J. Pers.*, **25**, 1956, 176–190.

MURRAY, H. The psychology of humor. *J. abnorm. soc. Psychol.*, **29**, 1934, 66–81.

REDLICH, F., LEVINE, J., & SOHLER, T. A Mirth Response Test: Preliminary report on a psychodiagnostic technique utilizing dynamics of humor, *Amer. J. Orthopsychiat.*, **21**, 1951, 717–734.

ROGERS, C., & DIAMOND, ROSALIND. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

WORCHEL, P. Adaptability screening of flying personnel: Development of a self-concept inventory for predicting maladjustment. *USAF Sch. Aviat. Med. Rep.*, 1957, Randolph AFB 56-62.

(Received August 6, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 271-275

THE DEVELOPMENT OF MEANING IN CHILDREN'S DESCRIPTIVE TERMS[1]

SUSAN M. ERVIN AND GARRETT FOSTER[2]
University of California, Berkeley

CHILDREN frequently confuse the names for physical dimensions. If two objects differ in size, they may say one is STRONGER than the other. If they differ in weight, one may be called BIGGER than the other. This is, of course, just what one would expect in the early stages of learning if size, weight, and strength are empirically correlated.

Osgood, Suci, and Tannenbaum (1957) have found correlations of a similar sort in examining the structure of meanings measured by semantic differential scales. Three factors have been found repeatedly—Evaluation, Potency, and Activity. Little attention has been given to the development or origin of these factors. Inspection of the scales defining each factor suggests that two conditions would create correlations. One is verbal conjunction; if whatever is said to be GOOD is also said to be FAIR the scales will be correlated through verbal associations. Secondly, "ecological covariation" (Brunswik, 1947) exists for certain sensory dimensions. We would thus expect BIG, HEAVY, and STRONG to be correlated for all cultural groups.

It is clear that adults can discriminate the sensory dimensions that children confuse verbally. On the semantic differential, however, they are normally asked to extend terms metaphorically, as in judging the size and weight of FREE PRESS or EDUCATION. Even on the semantic differential, the correlation of size and weight scales can be destroyed by inclusion of items like DIAMOND and MIST where a check with sense experience is possible and the usual trait correlation is reversed. To the extent that the semantic differential reflects covariation in experience of traits that are logically independent, we would expect that there would be an increase with age in denotative discrimination of the terms that are correlated on the differential. In the following study, age changes in children's verbal confusions are examined.

The more highly correlated two attributes are, the less probable are encounters with discrepant instances. It is useful to note three different variants on this situation. One category may constitute a subclass of a larger category. Thus, if 98% of a child's encounters with men, in which there is direct address or verbal reference to a man, involve his father, we would expect that the child might at first call all adult males DADDY. Discrepant instances would at first be too few for a differentiation to take place. If there were two adult males in the family such an extension would be unlikely. Thus Leopold (1939) noted that while his daughter called all men PAPA, women had individual names and there was no word for FRAU.

The subclass of a hierarchy of classes is actually an extreme instance of the second variant of correlation, a partial overlap of two classes. The degree of correlation or overlap should predict the probability of two terms being confused. Thus, communism and atheism might be confused by those unaware of discrepancies such as religious communist settlements. From a matrix showing the probability of being right in applying the term COMMUNIST to an atheist, we can see that the higher the correlation of the two terms, the greater the probability of being right, and the greater the likelihood that the two ideologies are called by one term. There are other features that enter into the failure to differentiate terms. One is the relative size of the two categories. If there are more atheists than communists, one is more likely to be wrong in calling an atheist a COMMUNIST than calling a communist an ATHEIST. The extreme case would be that in which the whole size of the class of communists is equal to the overlapping class. That is, for example, all communists are atheists but the reverse is not true. This case is identical with the one cited earlier of hierarchical classes.

---

[1] The authors acknowledge with gratitude the cooperation of Matt Griffeath, Principal, and the teachers of the Durant School in Oakland, California, in providing facilities and *S*s.

[2] Now at the University of Texas.

Other relevant factors are the perceptibility of an attribute or class, the consequences of correct and incorrect class discrimination (which may not be the same for the classes involved), the frequency of the terms in usage, and degree of logical independence.

If the dimensions of reference are continuous attributes rather than classes, then the relation may be described by a scatterplot rather than a matrix of frequencies. The same observations apply; the probability of being correct in saying that a bigger object is HEAVIER is a function of the attribute correlations.

In adult speakers of English, the differentiation of the attributes weight, size, and strength is such that if speakers can make appropriate tests, they are unlikely to say that the heavier of two like-sized objects is BIGGER. There are, however, situations in which the attribute extension in this simple physical case is appropriate. One is in the situation of prediction, where a value on one attribute only is known. Then it becomes useful to be able to predict the probable value on the other attribute. The second situation is one in which metaphor is exploited in verbal or pictorial communication, and one attribute may be used to suggest another.

Thus, we would expect that correlated attributes would appear in experienced speakers' usage, in situations of prediction and metaphorical extension, but not in denotation where the evidence for attribute discrimination is available.

In new learners, however, one term may apply to both attributes, which are not in fact discriminated, or both terms may appear as interchangeable synonyms for the two undiscriminated attributes.

In the following study two semantic differential factors, Evaluation and Potency, are presented as far as possible in conditions requiring denotative discrimination of three attribute expressions of each. A reduction with age is predicted with respect to the use of the wrong terms when a difference in only one attribute of the correlated set is present.

METHOD

Subjects. There were two groups of *S*s, a group of 16 male and 17 female first grade children, and a group of 18 male and 18 female sixth graders. Both groups were from the same school in a lower socioeconomic Negro district. To reduce variability, the extremely bright and extremely dull children were excluded, the criterion being the teacher's rating in the first grade, and deviations of 20 points from the norm on available IQ tests in the sixth grade.

Materials. In Part I of the experiment, materials were selected to vary successively three of the dimensions on Osgood's potency factor. These three dimensions—weight, strength, and size—had been found to have loadings of .62 with the rotated factor analysis, involving concepts rated against scales, and coordinates of 1.68, 1.81, and 1.76, respectively, on another analysis of scales judged against scales. These were the largest components of the factor in each case.

The objects used were: (*a*) opaque salt shakers identical but for weight, (*b*) opaque jars identical but for weight, (*c*) cork balls differing in size, (*d*) sterofoam balls differing in size, (*e*) a pair of insulated wires with the wire removed from the middle third of one, leaving it flexible, and (*f*) a dry sponge and a damp one matched in size.

In Part II drawings of a girl's face were used to represent three of four dimensions representing Osgood's evaluative factor. The pictured dimensions were CLEAN-DIRTY, HAPPY-SAD, and PRETTY-UGLY. A fourth was included in the questioning: GOOD-BAD. These had loadings, respectively, of .82, .76, .86 (BEAUTIFUL-UGLY), and .88, and coordinates of 2.38, 2.09, 2.40, 2.29 on the first factor.

Procedure. All of the *S*s were individually tested with the following questions:

> I would like to ask you some questions and you can give me the correct answers. . . OK? [For first graders:] It's kind of a game and lots of fun. First I'm going to ask about these objects. [2 objects contrasting in weight put in subject's hands.] Is one of these heavier and one lighter or are they both the same weight? [If says different] Which is heavier? Is one black and one white or are they both the same color? Is one bigger and one smaller or are they both the same size? [If says different] Which is bigger? Is one stronger and one weaker or are they both the same strength? [If says different] Which is stronger?

If the child failed to indicate the item that was heavier on the first question, he was eliminated from the rest of Part I. The second question, to which the answer was "the same" was to control set. A similar series of questions was asked about all the items, starting with the actual contrast as a screening question. One of the six possible key questions was omitted, concerning perception of weight in objects differing in size. If weight in the objects were controlled, the smaller object might be called heavier on the grounds of its scale weight. Because of the ambiguity of the term HEAVY applied to objects differing in size, the question was omitted.

In the analysis the percentages were computed over the whole set of responses, which included two for each question for each child because of the double set of materials. The *N*s used in the significance tests were for the actual number of children tested.

In Part II a similar procedure was followed, with the control question "Does one picture have red hair and one black hair or are they both the same?" The questions and pictures were rotated, with every fourth question a control question.

RESULTS

In Table 1 it can be seen that between 39 and 66% of the first-grade children offered contaminated responses for the various dimensions of the physical materials. In the sixth grade the proportion was reduced to a range between 20 and 44%. While none of the individual changes in percentage was significant, there was a reduction for every comparison, including the subgroups by sex. A sign test is significant at the .01 level. The dimension that changes the least is the response that the heavier object is STRONGER; this is also the statement of highest frequency at both ages for both sexes. Since strength is less evident than weight, it is possible that this particular inference would continue in adult *S*s. The inference that a bigger object is STRONGER is next in probability, supporting the notion that it is the inferred character of strength that is involved. Size is least often presented as a contaminated response, and it is also the most obvious.

It might be thought that the reduction in the sixth grade is due to a tendency to be more careful about differentiating at all, and thus to an increase in *same* responses. This was not the case. While the percentage of reversals was relatively low, it increased in the sixth grade. This increase occurred in stating that the heavier was SMALLER, that the stronger was SMALLER and LIGHTER. In the latter case these frequencies probably arise because a wet sponge is usually larger, and contains water. Only boys gave the last reversal, 22% of the boys in the sixth grade saying the stronger was LIGHTER, and no girls. Nine percent of the boys and 28% of the girls said the stronger was SMALLER. In the last case this was a larger proportion than those saying it was larger. It might be argued, then, that the reduction in the last two cases was due to sophistication with respect to one of the objects used.

On the study of the evaluative dimension using faces, almost no reversals occurred at either age. The age differences on Table 2 are striking. It may be noted that they do not occur markedly on three cases. Two of these refer to CLEAN. The frequency of children saying one child was CLEANER than the other for the other attributes was small at both ages. Clean may be said to be the most visible of the attributes. In fact it could be argued that it is the only one with a clear-cut physical criterion.

The third instance of lack of marked change was one in which the proportions were very high at both ages. The smiling face was said to be PRETTIER.

With the CLEANER dimension excluded, the range in the first grade is between 42 and 97% offering a contaminated response. The highest are those offering the smiling face as more GOOD and the cleaner face as PRETTIER. All of the first grade boys gave these responses. In the sixth grade the range is between 14 and 75% with the highest now being the smiling face which is seen as PRETTIER about as often as it was in the first grade. Thus, there is a shift in responses as with the physical attributes, but it is markedly different for the different attributes.

TABLE 1

PROPORTION ASCRIBING CORRELATED ATTRIBUTE DIFFERENCES TO OBJECTS

Stimulus Differences	Response	First Grade			Sixth Grade			
		Boys	Girls	Total	Boys	Girls	Total	Differences
Heavy	BIG	40.6	36.0[a]	38.6[a]	16.7	25.0	20.8	17.8
Strong	BIG	31.2	52.0	39.7	30.6	25.0	27.8	11.9
Strong	HEAVY	37.5	46.2	41.4	33.3	33.3	33.3	8.0
Big	STRONG	50.0	53.8	51.7	36.1	38.9	37.5	14.2
Heavy	STRONG	56.2	76.9	65.5	41.7	47.2	44.4	21.1
	N[b]	16	13	28	18	18	36	

[a] The number of cases in this cell is reduced by one due to loss of data.

[b] The *S*s reported were those who correctly differentiated the stimulus attribute for all three attributes on all materials.

TABLE 2
PROPORTION ASCRIBING CORRELATED ATTRIBUTE DIFFERENCES TO FACES

Stimulus Differences	Response	First Grade			Sixth Grade			
		Boys	Girls	Total	Boys	Girls	Total	Differences
Happy	CLEAN	12.5	5.9	9.1	0.0	11.1	5.6	3.5
Pretty	CLEAN	28.6	7.7	18.5	5.6	5.6	5.6	12.9
Clean	HAPPY	37.5	47.0	42.4	5.6	22.2	13.9	28.5*
Pretty	HAPPY	85.7	91.7	88.5	67.1	38.9	50.0	38.5*
Clean	PRETTY	100.0	82.4	90.9	50.0	50.0	50.0	40.9*
Happy	PRETTY	81.2	82.4	81.8	83.3	66.7	75.0	6.8
Clean	GOOD	50.0	70.6	60.6	27.8	33.3	30.6	30.0*
Happy	GOOD	100.0	94.1	97.0	77.8	55.6	66.7	30.3*
Pretty	GOOD	78.6	50.0	65.4	44.4	22.2	33.3	32.0*
	N[a]	16	17	33	18	18	36	

[a] Changes in N due to failure to name stimulus difference correctly: first-grade boys for PRETTY stimulus 14, girls 12; sixth-grade boys, 17, and girls, 16.

* Significant at the .01 level.

DISCUSSION

The slight rate of change with respect to discrimination of physical dimensions suggests the kind of learning to be expected where the criteria are most obvious.

With the personal attributes the findings are both more extreme and more uneven. It is clear that the only term of those used which designates a simple visible trait—CLEAN—is the one seldom offered to describe any other attribute. This finding agrees with the fact that BIG, the most obvious physical attribute, was less often used with the physical materials to describe other differences.

The other evaluative dimensions refer to more complex traits that are not entirely logically independent. PRETTY may also be said to designate a physical characteristic, but one that adults use both as a constant and a temporary trait, so that a clean smiling face might be deemed prettier than a dirty or frowning one. Thus, while the traits are discriminable they are not independent in the sense that the physical traits are.

HAPPY was used to designate a smiling face, but as children learn the use of the term they may find that it refers to a state of feeling only partially correlated with external evidence. Thus even the smiling–nonsmiling distinction might bear only an imperfect relation to the term HAPPY. One of the largest changes was in the use of HAPPY in describing the prettier girl, and the drop was most extreme in girls—a drop of 53%. The children were from the start, especially the boys, only moderately likely to describe the cleaner girl as HAPPIER.

One possible reason for a drop in the ascription of terms to a correlated difference is a change in metaphorical treatment of pictures. Occasionally children would refuse to say that both children were the same in hair color, but said that one had red and one black hair. The hair in the drawing was white, that is, not filled in in the black-and-white outline picture. Younger children, used to story-book imagination, may be less literal about what is on the page. While this might account for the increase in *same* responses with age, it does not predict the direction of the ascription, by the first graders, which was in no case in the opposite direction from that predicted.

Does the failure to differentiate on this test imply that the children use the terms as virtual synonyms for an undifferentiated referent? With respect to the physical dimensions, only one child was so extreme as to use the same terms interchangeably for all three attributes on all the materials. But if we examine the faces test and omit CLEAN which seems to fall out of the pattern, it appears that 62% of the children used HAPPY-GOOD-PRETTY synonymously, in the first grade.

We would like to argue that the factors that appear as clusters of correlated terms in the semantic differential studies of adults derive from empirical correlations of attributes. They could, of course, be linked purely by verbal associations, as in "He's a good clean player."

If this were the case, differentiation of reference might still be accompanied by semantic differential correlations, since many of the terms on the semantic differential can be applied to the "concept" only metaphorically. DEMOCRACY is clean, fragrant, and sweet only in the poetic sense.

While verbal associations may be one source of such dimensions, we are proposing that the history of concept development in the child provides another source. What remains as a connotative, metaphorical relationship in adults may in many cases start as denotative nondifferentiation. In a sense, the child might be said to acquire first a concept, for instance, of "big-strong-heavy . . ." in other words, a potency referent. The terms he applies to this referent may variously be BIG, STRONG, HEAVY. He may prefer one of these terms for people, another for boats, another for baseballs. Presumably he will only come to differentiate the terms and apply them appropriately to different stimulus dimensions when uncorrelated instances occur and he is corrected, or hears others differentiate the terms. By chance, the sample he selects may have a 100% correlation and he may not encounter errors immediately.

SUMMARY

The physical dimensions of size, weight, and strength are empirically correlated. If the correlation delays discrimination of these attributes as referents for descriptive terms, then younger children should more often use incorrect terms to describe differences between objects. The terms GOOD, PRETTY, CLEAN, and HAPPY should also be used as synonyms prior to differentiation.

A set of materials was prepared in which size, weight, and strength were independently varied in pairs of objects. First-grade children more often than sixth graders said that the pairs of objects differed on other dimensions in addition to the attribute actually contrasted. In a set of pictures of faces, over half of the youngest children treated GOOD, PRETTY, and HAPPY as interchangeable synonyms. The proportion dropped markedly with age. The more easily identified traits, such as the referents of BIG and CLEAN, were least often confused with other attributes.

The results are interpreted as showing that attributes which have metaphorical and connotative links in adult usage, may be denotatively confused at first. The factors found by Osgood on the semantic differential studies of verbal meaning may actually be the referents for several terms used as synonyms, prior to differentiation of finer distinctions between attributes.

REFERENCES

BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. California Press, 1947.

LEOPOLD, W. F. *The speech development of a bilingual child.* Evanston, Ill.: Northwestern Univer. Press, 1939–50.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

(Received August 7, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 276-286

A THERAPEUTIC APPROACH TO SCHIZOPHRENICS OF EXTREME PATHOLOGY:

AN OPERANT-INTERPERSONAL METHOD[1]

GERALD F. KING
Michigan State University
STEWART G. ARMITAGE
Veterans Administration Hospital, Battle Creek, Michigan
AND JOHN R. TILTON
System Development Corporation

One of the most severe forms of psychopathology is seen in cases culminating in a "vegetative" level of adjustment characterized by the patients being extremely withdrawn, anergic, and uncommunicative. Standard treatment procedures do not produce any appreciable benefits. It would seem that different, if not novel, therapeutic approaches need to be explored if changes are to be effected with this type of patient. This paper presents an evaluation of a therapeutic method that was devised for chronic schizophrenics with such extreme withdrawal symptoms.

Given the label "operant-interpersonal," the method of therapy that is proposed stems from recent work employing learning techniques with schizophrenics. Peters and Jenkins (1954) attempted to retrain chronic schizophrenics by using first simple psychomotor tasks and then increasing their complexity. To increase motivation, subcoma injections of insulin were given with fudge as a reward. Tilton (1956), employing instrumental motor and verbal techniques with even more debilitated patients, found that candy was an adequate reward in itself. His results revealed that mute patients responded better to the experimental sessions (mainly psychomotor in nature) than more verbal patients, while just the opposite trend was observed for "standard" verbal therapy, one of his control conditions. In a similar vein, Lindsley (1954) has discussed the possible extension of operant conditioning techniques to therapy.

Operant-Interpersonal Method: Rationale

The therapeutic goal is to elicit more interest in the environment and to promote more appropriate response patterns. Assumptions about the underlying nature of schizophrenic behavior are not reflected in the therapeutic procedure, as virtually no attention is given to the patient's previous experiences or pathology. Thus, in light of the common tools associated with psychotherapy (Shoben, 1949), the operant-interpersonal method might be considered an unsophisticated therapeutic approach. Be that as it may, the following considerations are offered as the basis for developing the present therapeutic procedure with *highly withdrawn* patients:

1. The therapeutic sequence should be built on responses that are in the repertory of the patient. Verbal communication is an inappropriate response medium because it has all but been abandoned by many patients as part of the withdrawal pattern. Ward observations reveal that although such patients show little verbal activity, they still shuffle about the ward, manipulate eating utensils, smoke cigarettes, and so forth. In short, motor responses are more intact and, hence, more usable.

2. The course of therapy should allow an easy transition from the simple to the more complex. An increase in complexity refers to the addition of verbal and interpersonal components to the therapeutic structure.

3. In the role of "friendly leader," the therapist

---

[1] Originally conceived by the senior author and David W. Merrell, this research was conducted at Veterans Administration Hospital, Battle Creek, Michigan, during 1956.

The authors are indebted to the following individuals for assistance: Carl R. Brown, being largely responsible for the design and construction of the apparatus; Norman Graff and David W. Merrell, for collecting portions of the data; M. Ray Denny and Norman R. F. Maier, for their interest and suggestions; the late Edmund D. Barry, for conducting the recreational therapy group; Terrence M. Allen, for his statistical consultation; and Frank Restle, for his advice on the preparation of this manuscript. Grateful acknowledgement is also due John Berghost and other personnel of the Medical and Nursing Services for their extensive cooperation.

should provide guidance in a social atmosphere that is conducive to learning. The therapist should also function as a decision maker, continually evaluating the patient's behavior and judging whether or not the patient is ready to cope with more complex situations.

METHOD

Measurements

The following instruments constitute the principal measuring devices that were employed. Others of a minor nature will be described in the analysis of the data.

Indices of status. In earlier work with deteriorated, uncommunicative patients, two of the authors devised the Extreme Mental Illness Schedule (EMIS), an interview procedure for assessing low levels of adjustment. A similar scale, which was apparently developed at approximately the same time, has been reported by Farina, Arenberg, and Guskin (1957). The EMIS consists of 22 variables, which are rated in an interview setting on well defined four- or five-point scales. The first 10 variables involve ratings on what might be called "simple interpersonal tasks" (e.g., response to handshake, response to offer of a chair, response to offer of a cigarette). The remaining variables were assembled from other rating scales, the emphasis being on nonverbal behavior (e.g., posture, psychomotor tempo, peculiar gestures). The sum of the ratings on all the scales yields a morbidity score, with higher scores representing greater severity of neuropsychiatric illness.

Using two staff psychologists to make independent EMIS ratings resulted in an interrater correlation of .95 ($N = 62$). With two different judges and different patients, a correlation of .92 was obtained for interrater reliability ($N = 50$). When the EMIS was readministered to 33 patients (all having achieved stable adjustments) after an interval of 8 weeks, the test-retest reliability was found to be .85. Several psychologists and psychiatrists were asked to rank small groups of patients on severity of illness. Converting the EMIS scores to ranks yielded validity coefficients (rho's) that ranged from .62 to .98. It was also possible to discriminate significantly between chronic schizophrenics with no privileges and chronic schizophrenics having limited privileges on the basis of EMIS morbidity scores.

Clinical status, as reflected by ward adjustment, was assessed with the Ward Observations Scale (WOS), the final 22 items of the Multidimensional Scale for Rating Psychiatric Patients (Lorr, 1953). In a pilot study, two attendants rated 20 chronic schizophrenics on the WOS, resulting in a correlation of .92 between the two independent morbidity scores. When a different attendant rated the same patients after an interval of 2 weeks, the test-retest correlation was .86.

A five-point scale for measuring amount of verbalization (AVS) was designed to be filled out at the end of the EMIS interview. The independent AVS ratings made by two staff psychologists correlated to the extent of .89 ($N = 62$). After 8 weeks, the test-retest reliability was .81 ($N = 31$).

Measures of clinical improvement. Difference scores obtained from successive ratings on the EMIS, WOS, and (to some extent) AVS provide measures of clinical improvement. In addition, the Clinical Improvement Scale (CIS) (King, Merrell, Lovinger, & Denny, 1957) was employed. The first section of the CIS consists of a simple five-point scale, in which the anchoring descriptions range from "has become worse" to "shows very marked improvement." In the event clinical improvement of any degree is indicated, the second section provides a check list which allows the rater to specify the nature of the behavioral changes. Previous studies (King, et al., 1957; Tilton, 1956) have indicated that ward personnel can make reliable assessments on the global scale of the CIS (first section). The following intercorrelations among the principal measures of clinical improvement were obtained ($N = 50$): EMIS vs. WOS, .51; EMIS vs. CIS, .62; and WOS vs. CIS, .45.

Subjects

Selection. All *S*s were drawn from a large locked ward which housed approximately 300 long-term, male psychotics. The following criteria were used in soliciting possible *S*s from the ward personnel: diagnosis of schizophrenia, under 45 years of age, low verbal and physical activity level, and no major change in behavior for at least two years. This procedure yielded 92 patients who were screened with interviews and then administered the EMIS and AVS. With the latter instrument, all patients were classified for level of verbalization using the following categories: mute or nearly mute, minimal verbalization, and below normal verbalization. Four groups of 12 *S*s each were formed so that each group contained four *S*s from each of the three verbalization levels. The *S*s of the four groups were individually matched on severity of illness (EMIS) and length of hospitalization.

Characteristics. The ages of the final 48 *S*s ranged from 25 to 42 years ($M = 33.8$, $SD = 4.4$), with length of hospitalization extending from 50 to 171 months ($M = 108.0$, $SD = 33.3$). Level of education was from 7 to 13 years ($M = 10.1$, $SD = 1.9$). The distribution of the schizophrenic subtypes was as follows: hebrephrenic, 18; undifferentiated, 14; catatonic, 11; and paranoid, 5. No significant differences existed among the four groups for these variables. All 48 *S*s had in the past undergone at least three of the following treatment procedures: electroshock therapy, insulin coma therapy, ataractic drug therapy, psychotherapy, and group therapy.

The Operant-Interpersonal Method

Apparatus. Section A of Figure 1 provides a diagram of the Multiple Operant Problem-Solving Apparatus (MOPSA), which was constructed to put the operant-interpersonal method into practice. A modification of the apparatus used by Skinner, Solomon, and Lindsley (1955–56), MOPSA was built into an 8 by 8 ft. panel erected across the alcove of a large room. Projecting from the middle of the panel, 35 in. from the floor, were three wooden levers (6 in. in length) arranged in a row and spaced 5 in. apart. Below the outer levers were slots opening into trays, which were used for dispensing

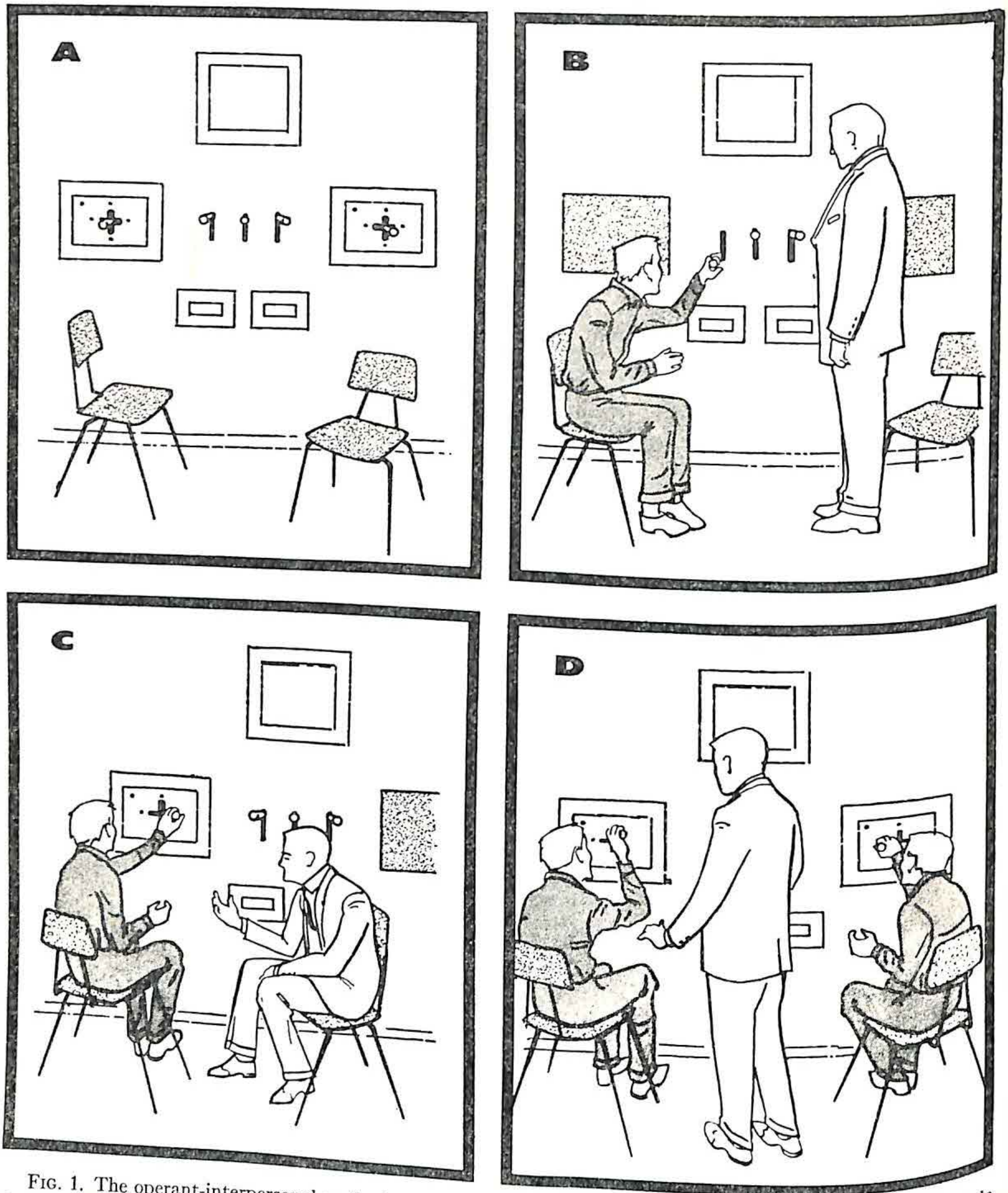

FIG. 1. The operant-interpersonal method: apparatus and illustrations of the therapeutic phases. (A. Multiple Operant Problem-Solving Apparatus—MOPSA, B. Simple Operant Behavior—Phase I, C. Simple and Complex Problem-Solving—Phase II, D. Cooperative Problem-Solving—Phase III)

rewards of candy and cigarettes. A 14 by 14 in. screen for projecting slides was located above the middle lever protected by a Plexiglas cover. Each lever could be depressed 3 in. against a force of 3 lbs., as regulated by an extension spring. In order to obtain rewards (candy, pictures, cigarettes), it was necessary for *S* to depress the appropriate lever until there was a "click" followed by the ringing of a chime.

Lateral to each of the outer levers by 8 in. were two 11 by 13 in. aluminum plates, which could be concealed by covers. A cross, into which a lever could be inserted, had been cut in the center of each plate. Counterbalanced by just enough tension to return and maintain them at the cross intersections, the levers could be moved freely in the two planes of the crosses. If *S* moved the lever of either cross from the position of rest to the end of the correct crossarm, a green light at the end of the arm became lit. Through a system of relays, problems could be introduced requiring from one to four movements (lights), along with various patterns of the "right-left-up-down" sequence. Movements had to be executed in a prescribed order so that an error canceled any previously correct responses, requiring that the entire sequence be started again. In addition to the green lights at the crossarms, an amber light located at an upper corner of each plate flashed when a problem was solved. By changing the relay system, new problems could be introduced in a few seconds. As the two crosses could be operated independently or together, problems requiring cooperation between *S*s could be employed.

Temporary screens enclosing an 8 by 10 ft. area were erected in front of MOPSA to reduce distraction and to restrict the movements of *S*. During the initial stages of therapy, the only furnishings in this area were two chairs. Behind the panel and not visible to *S*, a therapy-assistant operated the projector (scenic slides), released the candy (Hershey chocolate kisses) and cigarettes (Lucky Strikes), operated the relay system, and recorded the readings of the counters attached to the operant levers.

Therapeutic procedure. The 12 *S*s undergoing operant-interpersonal therapy were seen three times a week for 3½ months (15 weeks). Therapeutic sessions lasted from 20 to 30 minutes. In adapting the operant-interpersonal method to the individual *S*, variations were bound to occur. However, the therapeutic procedure can be divided into the following three phases which applied to all *S*s: Phase I, Simple Operant Behavior; Phase II, Simple and Complex Problem-Solving; and Phase III, Cooperative Problem-Solving.

At the beginning of Phase I, each *S* was led into the therapy room and seated in front of MOPSA. After directing *S*'s attention to the three operant levers, the therapist briefly described their operation. The explanation was followed by a demonstration in which the therapist manipulated each of the levers, pausing to present the candy and cigarettes to *S* and to call his attention to the picture on the screen. The *S* was then urged to depress one of the levers. In a majority of the cases, it was necessary to repeat the demonstration. In a few cases, as a last resort, the therapist placed *S*'s hand on the lever and then supplied the necessary force for depressing the lever. With one *S*, the therapist initially unwrapped the candy reward and placed it in the *S*'s mouth. If *S* successfuly manipulated a lever, his performance was greeted with a brief but definite acknowledgment (e.g., "good," "very good," "good work," etc.). Section B of Figure 1 gives an illustration of this simple operant behavior. Verbalization by the therapist was held to a minimum, and what was said was mainly directed toward the operation of MOPSA.

Although there was marked variability among *S*s, all eventually began to respond voluntarily. When the individual *S* reached this point, frustration was introduced in the form of nonreinforced trials. When any of the *S*s stopped pressing the levers (extinction) after nonrewarded trials, special attention (instructions and demonstrations) was given by the therapist until *S* was willing and able to respond to various conditions of nonreinforcement. Phase I lasted approximately 4 weeks.

In Phase II, one of the aluminum plates was uncovered, and a lever was inserted into the cross. Beginning with the simplest type of problem solving, the therapist demonstrated the one-light problem, which required the lever to be moved into a single crossarm. He showed that only one of the four possible movements turned on a green light (at the end of the crossarm) and subsequently the amber light (upper corner of the plate). After making the correct response on the cross, the therapist depressed one of the operant levers and called *S*'s attention to the resulting reward. When *S* could satisfactorily solve a variety of one-light problems, he was gradually introduced to problems requiring more movements. Eventually all *S*s reached at least the level of two-light problems, while five of the *S*s consistently solved four-light problems. This problem solving behavior is illustrated in Section C of Figure 1.

After approximately 3 weeks, verbalization was integrated into the problem-solving process. For example, *S*, after solving a two-light problem several times, might be stopped after making the first movement and asked, "Now where do you go next?" If the therapist failed to elicit a verbal response from a mute *S*, he would ask the *S* to point to the next movement. Another procedure was to ask *S*, after he had solved a problem and received a reward, to describe all the movements required in the problem. When *S* was confronted with a new problem, he might be asked, "What is going to be your first move?" At this time, the therapist also encouraged more verbal interaction at the beginning and close of sessions, with topics not directly related to MOPSA being discussed (e.g., "What did you have for breakfast this morning?"). Phase II lasted about 6 weeks.

The second cross was uncovered at the outset of Phase III, which was devoted to cooperative problem solving. Seated in front of the second cross, the therapist informed *S* that they would have to work as a team in order to obtain rewards. In a simple two-light problem, the therapist made the first movement and *S* the second, as well as vice versa. Quality of performance seemed to drop sharply in this interpersonal context, as most *S*s experienced difficulty in coordinating their efforts with another individual. As progress did occur, the therapist and *S* proceeded to three- and four-light problems in which all possible therapist-subject orders of responding were required. As in the latter stages of Phase II, verbal interaction was emphasized.

When all *S*s had considerable experience working with the therapist, two *S*s functioned as a team. This feature of Phase III is shown in Section D of Figure 1. Then, two additional *S*s entered the therapeutic situation, standing behind and observing the cooperating *S*s before being called on to give either motor or verbal solutions. Stand-by chairs were pulled into the therapeutic area, and eventually as many as six *S*s attended a single session. In this way, it was possible for every *S* to cooperate with every other *S* within a short period of time. Occasionally, the therapist again functioned with more advanced *S*s and purposely made mistakes, placing the burden of making the corrections on the *S*s. A by-product of this procedure was the introduction of some mild humor into this situation. Problems were also devised in which two *S*s, after performing successfully, were required to communicate the solution to naive *S*s (nonobservers). Phase III was terminated after about 5 weeks.

Control Groups

The following provides brief descriptions of the control groups which were conducted concurrently with the operant-interpersonal group.

Verbal therapy. The 12 *S*s undergoing verbal therapy were given the same amount of time as the operant-interpersonal *S*s, three sessions of 20 to 30 minutes per week. Therapy was adapted to the individual *S*. The therapist made every effort to establish verbal communication with them on whatever basis was possible. Depending on the *S*, a variety of topics were explored, ranging from routine daily events to *S*'s feelings and concerns. With mute *S*s, the sessions consisted of monologues by the therapist, punctuated by hopeful pauses. When operant-interpersonal therapy began using more than one *S*, a similar procedure was followed in verbal therapy. Hence, in the latter stages verbal therapy became group verbal therapy.

Recreational therapy. The 12 *S*s in recreational therapy participated in a variety of activities, usually within a larger group of patients. Time spent in this form of therapy ranged from 3 to 5 hours per week.[2] The following indicates the nature of some of the activities: medicine ball manipulation, shuffleboard, rhythm band, singing, social games (e.g., hokey-pokey), dancing with volunteer workers. For the purpose of the present research, the *S*s were given more individual attention than they ordinarily would have received.

No therapy. Although the 12 *S*s in this group were not given any special treatment, they were allowed to continue to participate in the ward activity schedule (e.g., occupational therapy, corrective therapy), as did the other 36 *S*s on a more limited basis. Heretofore, the program had produced no discernible change in the patients of the ward. Except when it was not possible (e.g., swimming), most of the *S*s slept during the activities.

Procedure

All 48 *S*s were transferred to a single dayroom in which they were the sole occupants. A specially selected attendant was transferred from night duty to function as an attendant-observer on the dayroom for several hours each day.

The *S*s had been administered the EMIS and AVS by one of the authors prior to their being assigned to groups. Teams of attendants and nurses rated the *S*s on the WOS. A staff psychologist not previously associated with the research readministered the EMIS and AVS at the end of the project (15 weeks later). The attendant-observer rated the *S*s again on the WOS and filled out the CIS. Although the attendant-observer was aware that the *S*s were participating in a special project, he was not acquainted with the nature of the various groups or their memberships.

The *S*s of the operant-interpersonal and the verbal therapy groups were divided among two therapists (two of the authors) who were assigned six *S*s from each group. Each of the four subgroups were represented by two *S*s from each level of verbalization. The decision to use the two available therapists in this manner might be questioned since the same therapists saw the *S*s in both the operant-interpersonal group and one of the control groups. Although the personality characteristics of the therapists would be held constant by this procedure, there was the possibility of differential effort or interest. To combat any possible bias, the therapists attempted to consider each *S* on an individual basis and to utilize all possible resources within the requirements of the research design. Detailed notes were kept for all therapeutic contacts. The therapists periodically monitored each other's sessions as hidden observers. The *S*s receiving recreational therapy were assigned to an experienced activity-therapist who was particularly interested in chronic schizophrenics.

RESULTS

Since the *S*s from the operant-interpersonal and verbal therapy groups were divided among two therapists, a preliminary analysis was concerned with any differences in the two groups attributable to the therapists. Improvement scores derived from the EMIS and WOS were each treated by analysis of variance. As none of the relevant *F* ratios were statistically significant at the .05 level, the null hypothesis was accepted, i.e., no differences produced by the therapists.

Differential Clinical Improvement

Table 1 gives the results of the analysis of variance for the EMIS improvement scores. As can be seen, a significant difference among the therapeutic groups was indicated by an *F* ratio of 7.313 ($p < .001$). Initial status (severity of illness and/or length of hospitalization) also produced a significant *F* ratio ($p < .05$) using

TABLE 1

ANALYSIS OF VARIANCE OF EMIS IMPROVEMENT SCORES

Source	*df*	*MS*	*F*
Therapeutic Groups (T)	3	58.965	7.313**
Level of Verbalization (V)	2	22.312	1.057
Initial Status within V (I)	9	21.118	2.619*
T × V	6	3.924	0.487
T × I	27	8.063	
Total	47		

* $p < .05$.
** $p < .001$.

---

[2] The group receiving recreational therapy was a somewhat unusual control group. Due to a misunderstanding, this group was seen more than twice as much as the operant-interpersonal and verbal therapy groups.

T × I as the error term, which is not entirely appropriate in this case. Contrary to Tilton's (1956) findings, no significant difference was found for level of verbalization, although it should be pointed out that only *S*s relatively low in verbalization were used in the present research. The operant-interpersonal group showed the most EMIS improvement, with the control groups being in the following descending order: recreational therapy, no therapy, verbal therapy. It was considered appropriate, and not a post hoc analysis, to compute *t* tests between the operant-interpersonal group and the means of each of the control groups. The results, using two-tailed tests, revealed that the operant-interpersonal group improved significantly more than the recreational therapy group at the .02 level and both the no therapy and verbal therapy groups at beyond the .01 level.

When the WOS improvement scores were treated by analysis of variance, only the difference among the therapeutic groups was significant ($F = 4.652$, $p < .01$). In fairly close agreement with the previous analysis, the descending order of the means of the groups on improvement was as follows: operant-interpersonal method, recreational therapy, verbal therapy, no therapy. While *t* tests revealed that the difference between the operant-interpersonal and recreational therapy groups was not statistically significant, the operant-interpersonal group showed a significantly higher level of improvement than the verbal therapy and no therapy groups, both at the .01 level. The full Lorr scale consists of 12 factors or dimensions derived from factor analysis, with the WOS contributing at least three items to each of the following factors: activity level, compliance vs. resistiveness, submissiveness vs. belligerence, and withdrawal. Using scores based on each of the four factors, differences among the therapeutic groups were further explored with analysis of variance. Only improvement in the area of withdrawal yielded a significant difference ($F = 3.827$, $p < .05$).

TABLE 2

COMPARISON OF THE THERAPEUTIC GROUPS: DISTRIBUTION OF THE SUBJECTS ON THE CLINICAL IMPROVEMENT SCALE (CIS)

CIS[a] Ratings	Groups			
	Operant-Interpersonal	Verbal Therapy	Recreational Therapy	No Therapy
Has become worse	0	1	0	0
Essentially no change	4	10	8	10
Shows minor improvement	3	1	2	2
Shows considerable improvement	5	0	2	0

[a] The rating "shows very marked improvement" is not included in the table since no *S* was rated in this category.

TABLE 3

COMPARISON OF THE THERAPEUTIC GROUPS ON CHANGES IN LEVEL OF VERBALIZATION

Changes in Level of Verbalization (AVS)	Groups[a]			
	Operant-Interpersonal	Verbal Therapy	Recreational Therapy	No Therapy
Decreased	0	6 (7)	2	1
Unchanged	5 (5)	6 (5)	8	9
Increased	7 (7)	0	2	2

[a] The numbers in parentheses for the operant-interpersonal and verbal therapy groups represent appraisals of changes in level of verbalization obtained from the notes kept by the therapists.

Further support for the greater effectiveness of operant-interpersonal therapy can be found in Table 2, which presents a comparison of the therapeutic groups on the CIS. More *S*s in the operant-interpersonal group (eight) exhibited some degree of clinical improvement than *S*s in any of the control groups. Ranking the groups on amount of improvement from Table 2 results in the identical order as that obtained for the EMIS scores. The operant-interpersonal group was compared with each of the control groups with the Fisher exact test by using the dichotomy "improvement vs. no improvement" and constructing fourfold tables. With this analysis, the operant-interpersonal group was significantly better than the verbal therapy ($p = .009$) and no therapy ($p = .019$) groups; the trend in favor of the operant-interpersonal over the recreational therapy group was not significant ($p = .220$). Although nothing is known about its reliability, the CIS checklist (second section) provides information on the qualitative aspects of clinical improvement. An inspection of the checklist indicated that the nature of the clinical changes was quite similar for the improved *S*s of the operant-interpersonal and control groups. The largest discrepancies occurred on the checklist items "improved verbal communication," where proportionately more improved *S*s of the operant-interpersonal group were checked, and "less tension," where the incidence of controls was higher.

Changes in level of verbalization, as obtained from successive administrations of the AVS, are given in Table 3. The operant-interpersonal group was again superior. More *S*s in this group showed an increase in verbalization, and it was the only group in which no *S* dropped in amount of verbalization. A striking feature revealed by Table 3 is that at least six *S*s in the verbal therapy group showed a decrease in verbalization. When the dichotomy "increase vs. no increase" was used for the verbalization changes, the Fisher exact test revealed that the operant-interpersonal group was significantly different from the verbal therapy group ($p = .005$). The comparisons of the operant-interpersonal group with the recreational therapy and no therapy groups both approached significance ($p = .089$). These tests of significance are conservative ones, since the classification for change does not take into consideration decreases in verbalization.

During the last 3 weeks of the therapeutic sessions, one of the authors began attending as an observer the meetings of the ward personnel in which various patients were discussed and evaluated. At this time, patients were being considered for transfer to a small "privilege" section which had been opened on the ward. The observer recorded the names of all patients receiving favorable consideration for transfer. None of the ward personnel participating in the meetings knew what treatment procedures were being given the patients who were serving as *S*s in the research. Listed according to group memberships, the number of *S*s recommended for transfer was as follows: 5 (operant-interpersonal), 0 (verbal therapy), 2 (recreational therapy), 0 (no therapy). Actually, only one *S*, a member of the operant-interpersonal group, was transferred to the open section within 4 weeks after termination of the therapeutic sessions.

Subsequent evaluations. Some data on the status of the therapeutic groups were collected 6 months after the therapeutic sessions were discontinued. At this time, a team of ward personnel rated the *S*s of the operant-interpersonal, verbal therapy, and no therapy groups on the WOS. The recreational therapy group was not included since it, as an exception, was still receiving therapeutic attention. Treatment of the WOS morbidity scores by analysis of variance yielded a significant difference among the three therapeutic groups ($F = 4.239$, $p < .05$), with the operant-interpersonal group having the least pathology. The operant-interpersonal group was the only one showing a significant improvement over the initial WOS rating ($t = 2.664$, 11 *df*, $p < .05$). The transferred *S* from the operant-interpersonal group was still functioning on the open section. However, no other *S*s had been transferred during this time interval.

Other Comparative Data

When the screening interview was conducted, none of the *S*s, as best as it could be determined, indicated any interest in leaving the ward for any purpose. During the thirteenth week of the therapeutic sessions, a staff psychologist not previously associated with the research briefly interviewed all *S*s of the operant-interpersonal, verbal therapy, and no therapy groups. Each *S* was rated for motivation to leave the ward (either to an open ward or out of the hospital) on the following scale: 1—no interest, 2—some interest, 3—definite interest. It can be seen in Table 4 that more *S*s in the operant-interpersonal group displayed some degree of interest in leaving the ward than the *S*s in the two control groups. When this measure of motivation was divided into "interest vs. no interest," the Fisher exact test indicated that these differences were significant ($p = .019$) when the operant-interpersonal group was compared with each of the control groups.

As was noted earlier, the *S*s of the no therapy

TABLE 4

COMPARISON OF THE THERAPEUTIC GROUPS ON DEVELOPMENT OF MOTIVATION TO LEAVE THE WARD

Ratings on Motivation to Leave the Ward	Groups		
	Operant-Inter-personal	Verbal Therapy	No Therapy
No interest	4	10	10
Some interest	5	1	2
Definite interest	3	1	0

group continued to participate in the ward activity schedule. Although less time was available due to the various therapeutic procedures, the other *S*s also attended these activities. The one activity in which all *S*s participated was occupational therapy. At the close of the therapeutic sessions, the occupational therapist was given a list of all *S*s and asked to indicate any who had displayed "increased interest or more constructive behavior" in occupational therapy during the preceding 2 months. The resulting distribution of *S*s rated as showing some positive change was as follows: 4 (operant-interpersonal), 0 (verbal therapy), 1 (recreational therapy), 0 (no therapy).

In the initial evaluation of the *S*s, it was found that each therapeutic group contained from one to three *S*s classified as enuretic. At the close of the therapeutic sessions, only one *S* was designated as improved by the night attendants, and this *S* was a member of the operant-interpersonal group.

If the therapy notes compiled by the therapists are used, another comparison between the operant-interpersonal and verbal therapy groups can be made. The notes indicated that during the first 3 weeks two *S*s from each group showed resistance, i.e., resistance to leaving the ward for the therapeutic sessions. No *S* from the operant-interpersonal group behaved in this manner during the last 3 weeks, but five from the verbal therapy group were described as resistant.

Clinical Observations

The progress shown by the *S*s in the operant-interpersonal group was readily discernible to the authors and other daily observers. To give an example, there was the day during the eleventh week of therapy that a long-term mute *S* surprised the therapist by greeting him with, "Good morning. Got a match?" The authors were particularly impressed with the interest value that the therapeutic tasks held for the *S*s. More than a month after therapy had been terminated, an apparently puzzled attendant contacted one of the authors to inquire about the "lights," as two former *S*s were constantly asking him when they could resume this activity. Observations of the *S*s in the verbal therapy group paralleled the formal analysis. A number of *S*s in this group obviously became worse. Strong negativism toward the therapeutic sessions was common, and practically all of the eight verbal *S*s responded less, several becoming mute.

Interrelationships among Subject, Therapeutic, and Outcome Variables

The interrelationships among the following types of variables were explored with the 12 *S*s of the operant-interpersonal group: subject, therapeutic, and outcome. The subject variables were age, length of hospitalization, severity of illness, and level of prepsychotic adjustment. Severity of illness was indicated by a composite of the EMIS and the WOS. Using a scale derived partly from certain prognostic criteria (Kantor, Wallner, & Winder, 1953; Phillips, 1953), one of the authors ranked the *S*s on level of prepsychotic adjustment on the basis of case history information. Previous research (King, 1958) has shown a relatively high interrater reliability for ranking a similar dimension from case history material. The therapeutic notes were used by two of the authors who collaborated in ranking the *S*s on the following therapeutic variables: operant rate, level of problem solving, and task improvement. Rate of operant response (number of responses) was based on performance during the first 4 weeks of therapy. Level of problem solving pertained to the ability of the *S*s to solve the light-sequence problems of Phase II. Task improvement reflected the ranking of the *S*s on overall improvement in coping with the task requirements throughout the course of therapy. Outcome was represented by the relative standings of the *S*s on clinical improvement, as indicated by a composite of the improvement scores from the EMIS, WOS, and CIS.

The rank-order intercorrelations among the various subject, therapeutic, and outcome variables can be examined in Table 5. Of the subject variables, only severity of illness was related to any of the therapeutic variables, the rho's with operant rate and level of problem solving being significant at the .05 and .02 levels, respectively. The negative correlation between severity of illness and operant rate coincides with the results reported by Mednick and Lindsley (1958) with chronic patients. Although there was a trend for length of hospi-

TABLE 5

RANK-ORDER INTERCORRELATIONS AMONG SUBJECT, THERAPEUTIC, AND OUTCOME VARIABLES ($N = 12$)

Variables	2	3	4	5	6	7	8
1 Age	.804****	−.139	.028	−.070	.028	−.244	−.434
2 Length of Hospitalization		.189	.140	−.301	−.322	−.378	−.545*
3 Severity of Illness			.105	−.601**	−.685***	−.566*	−.266
4 Level of Prepsychotic Adjustment				.119	.098	.091	−.580**
5 Operant Rate					.699***	.615**	.203
6 Level of Problem Solving						.769****	.294
7 Task Improvement							.594**
8 Clinical Improvement							

Values of p (two-tailed tests):
* $p < .10$.
** $p < .05$.
*** $p < .02$.
**** $p < .01$.

talization to be negatively correlated with outcome ($p < .10$), the one subject variable significantly correlated with clinical improvement was level of prepsychotic adjustment ($p < .05$). Since it was negative, this latter correlation would appear to be inconsistent with previous findings which have shown that level of prepsychotic adjustment, as a reflection of more reactive than process characteristics, was positively related to prognosis. A possible explanation lies in viewing this finding in the light of an analysis presented by Zubin and Windle (1954), who found that certain prognostic indices were positively related to favorable outcome in acute psychotics but negatively related for chronic psychotics (the status of the present *S*s). All the therapeutic variables were positively intercorrelated at the .05 level or better, and each correlated positively with outcome. However, only the correlation between task improvement and clinical improvement was significant ($p < .05$). This correlation is compatible with the important assumption that therapeutic gains generalize to behavior outside of therapy. As in most post-mortem discussions, reasonable interpretations on the whole can be posited for the obtained correlations.

DISCUSSION

In giving fairly clear support for the therapeutic effectiveness of the operant-interpersonal method, the results concur with those reported by Peters and Jenkins (1954) and Tilton (1956) in their evaluations of similar therapeutic methods with chronic schizophrenics. There are few basic differences in the therapeutic orientations employed in these studies. While in the present study the analysis of therapeutic gains was based on measures tapping more facets of clinical improvement, the principal difference between this and the previous studies seems to be that the patients selected for therapy, in addition to being more debilitated, represented a more homogeneous group. The selection of patients was guided by the notion that the operant-interpersonal method was especially applicable to extremely withdrawn schizophrenics. It appeared that restricting the patients to the aforesaid type sharpened the differences between the operant-interpersonal method and the control procedures. For example, evidence was obtained indicating that control patients undergoing verbal therapy actually experienced some adverse effects. This finding is compatible with one of the underlying assumptions of the operant-interpersonal method, expressible as follows: in conducting therapy with patients of this nature, verbal communication should not be emphasized, at least, as an initial procedure.

The present study does not throw any appreciable light on what processes or components, among the many in the operant-interpersonal method, are crucial ones in promoting clinical improvement. How much does the relationship that develops between therapist and patient contribute to success? Without getting into reinforcement theory, one might wonder about the operation of the rewards. For example, is the patient more responsive to social reinforcement ("good") or to more tangible rewards (e.g., candy)? Actually, the different features of the method pose some problems of more general interest, as attested by the following question: does the introduction of an interpersonal component result in a qualitative change in the therapeutic setting, or does it merely represent increased complexity? In regard to the latter issue, there is some evidence indicating that, instead of or in addition to increased complexity, interpersonal components represent a separate dimension of withdrawal, interpersonal-withdrawal as opposed to thing-withdrawal (King, 1956).

The practical aspects of the findings bring forward other matters of interpretation. The judgments of the ward administrators led to only one patient of the operant-interpersonal group being transferred to an open ward, not an imposing gain. It is not difficult to think of a reasonable defense for this finding. Since the patients were only seen for approximately 22 hours, there is the possibility of additional clinical improvement with a longer period of therapy. Further sessions could be devoted to extending the previous therapeutic pattern, i.e., increasing the complexity of the therapeutic environment. The therapeutic tasks would gradually require more complex discriminations and more conceptual solutions as the problems at the same time became more and more embedded in a verbal and interpersonal context.

On the other hand, there is the possibility that the obtained degree of improvement represents all that can be expected from the operant-interpersonal method. Such an outcome would be compatible with a commonly accepted orientation which implies that a necessary condition for significant gains in *functional* disorders is therapy that penetrates beneath symptoms. The authors are willing to entertain, are even inclined toward, the possibility of limited further improvement, but for different reasons. For one thing, there is the question of whether the patients involved, instead of being clearly functional, do not have contributing organic factors (to be discussed in the next paragraph). In any event, if the results are accepted as they stand, the practical value of the operant-interpersonal method becomes less evident. No ready answer can be offered for the question "What comes next?" The authors resist the temptation to resort to the rationale so frequently associated with the newer psychopharmacological drugs: that the patients, while not "cured," are made accessible to psychotherapy. Whether or not organicity is a factor is one matter, but independently there are doubts about the extent to which psychotherapy can be applied. There is no convincing evidence that psychotherapy is a feasible method for *chronic* psychotics at any stage of illness.

The operant-interpersonal method might be given a final scrutiny in the light of the type of patients for which it was devised. In classifying such patients as being in the terminal stages of process schizophrenia, many clinicians tend to suspect, or even posit, the presence of organic factors either in their etiology or, at least, current functioning. Some evidence can be cited to support this notion (Brackbill, 1956). If this point of view is accepted for the sake of discussion, then psychological therapies for this class of patients would be peripheral approaches, ones that could only operate within certain "neurophysiological limits." Although beneficial results from purely psychological therapies would not be precluded, the implication is that therapeutic gains would be confined to symptomatic ones. Within the available limits, however, specificity of treatment might well be related to the amount of clinical improvement. According to this outlook, the operant-interpersonal method, by focusing on specific symptomatology (withdrawal behavior), becomes a reasonable *psychological* approach. This line of thought is similar to that offered by Jones (1956) in his discussion of therapeutic procedures for "constitutional" disorders.

SUMMARY

The operant-interpersonal method was developed as a therapeutic procedure for chronic schizophrenics with pronounced withdrawal symptomatology. In order to put the method into practice, the Multiple Operant Problem-Solving Apparatus was constructed. Emphasizing motor behavior, the operant-interpersonal method progressed from the simple to the complex. Initially, the patients undergoing this procedure on an individual basis made simple operant responses in the presence of a therapist for rewards of candy, cigarettes, and colored slides. More complex psychomotor, verbal, and interpersonal components were systematically incorporated into the procedure in accordance

with the patient's progress. At its maximum complexity, the therapeutic environment required each patient to communicate with other patients and to enter into cooperative relationships in order to solve problems.

Operant-interpersonal therapy was given to 12 patients, while an equal number of patients concurrently underwent each of the following control conditions: verbal therapy, recreational therapy, no therapy. For the patients in the operant-interpersonal and verbal therapy groups, therapeutic sessions of 20 to 30 minutes were scheduled three times per week for 15 weeks.

The operant-interpersonal method was more effective than all the control methods in promoting clinical improvement, based both on ward observations and interview assessments. Comparisons on the following variables also yielded differences in favor of the operant-interpersonal method: level of verbalization, motivation to leave the ward, resistance to therapy, more interest in occupational therapy, decreased enuresis, and transfers to better wards. The patients undergoing verbal therapy actually became worse in some ways (e.g., verbal withdrawal). Some comparative data were also collected 6 months after the termination of the therapeutic procedures. Several frames of reference, both theoretical and practical, were used in interpreting the results.

REFERENCES

BRACKBILL, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, **53**, 210–226.

FARINA, A., ARENBERG, D., & GUSKIN, S. A scale for measuring minimal social behavior. *J. consult. Psychol.*, 1957, **21**, 265–268.

JONES, H. G. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 414–419.

KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.

KING, G. F. Withdrawal as a dimension of schizophrenia: An exploratory study. *J. clin. Psychol.*, 1956, **12**, 373–375.

KING, G. F. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *J. abnorm. soc. Psychol.*, 1958, **56**, 160–164.

KING, G. F., MERRELL, D. W., LOVINGER, E., & DENNY M. R. Operant motor behavior in acute schizophrenics. *J. Pers.*, 1957, **25**, 317–326.

LINDSLEY, O. R. Discussant for B. F. Skinner, A new method for the experimental analysis of the behavior of psychotic patients. *J. nerv. ment. Dis.*, 1954, **120**, 403–406.

LORR, M. Multidimensional scale for rating psychiatric patients: Hospital form. *VA tech. Bull.*, 1953, No. 10-507.

MEDNICK, MARTHA T., & Lindsley, O. R. Some clinical correlates of operant behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 13–16.

PETERS, H. N., & JENKINS, R. L. Improvement of chronic schizophrenic patients with guided problem-solving, motivated by hunger. *Psychiat. Quart. Suppl.*, 1954, **28**, 84–101.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

SHOBEN, E. J., JR. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, **46**, 366–392.

SKINNER, B. F., SOLOMON, H. C., & LINDSLEY, O. R. New techniques of analysis of psychotic behavior. *ONR tech. Rep.*, 1955–56, No. 3. (Contract N5-ori-07662)

TILTON, J. R. The use of instrumental motor and verbal learning techniques in the treatment of chronic schizophrenics. Unpublished doctoral dissertation, Michigan State Univer., 1956.

ZUBIN, J., & WINDLE, C. Psychological prognosis in mental disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 272–281.

(Received April 13, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 287-291

CASE REPORT

A PSYCHOLOGICAL CASE STUDY OF DEATH DURING PREGNANCY[1]

ANTHONY DAVIDS AND MORTON SILVERMAN
Brown University and Emma Pendleton Bradley Hospital

In the course of a collaborative project investigating perinatal factors in child development, we have studied more than 100 pregnant women. In addition to exhaustive medical observations and obstetrical recordings on these patients, the women have been administered a comprehensive battery of psychological tests. The aim of the psychological assessment is to add to knowledge about emotional factors in pregnancy, and to relate the findings to medical data recorded in the delivery room. This collaborative investigation is planned as a long-range research program, and eventually we hope to relate emotional factors in the mother, during and shortly after pregnancy, to the subsequent intellectual and emotional development of the offspring. The emphasis in this overall research program is on well-controlled experimental design, large samples of patients, and statistical analyses of data. However, it seems worthwhile to report the findings derived from this individual case, which, fortunately, represents a rare occurrence in research contexts.

BRIEF CASE HISTORY

The patient was a 39-year-old white woman of the Protestant faith. She had been married 8 years to a man 11 years her senior and had a 6-year-old son. Immediately following delivery of the first child, the patient underwent cholecystectomy (removal of the gall bladder). Before her discharge from the hospital, a metastasizing carcinoid, originating in her duodenum and involving her pancreas, was diagnosed. In the course of major surgery necessitated by this cancerous condition, the pancreas and spleen were removed, and a surgical hitch-up between the small intestine and the stomach was performed. Removal of the pancreas resulted in diabetes and demanded continuous insulin treatment. Shortly after the operation, a subphrenic abscess developed, requiring further surgery to permit drainage. Shortly thereafter, the patient developed intestinal obstructions requiring reoperation. Then a postoperative hernia developed, with separation to the level of the skin, necessitating further major surgery to insert a tantalum screen in the wall of the upper abdomen. In addition to these abdominal difficulties, the patient had a history of chronic otitis media of unknown onset.

After the stomach surgery, the patient did not menstruate for many months. She had been strongly advised not to become pregnant, and when her menstrual cycles began again, she used controceptives. When she first learned that she was again pregnant in August 1959, she was quite upset, but then decided that she wanted to have the baby. Her surgeon advised against this and recommended a therapeutic interruption or legal abortion. She was referred to an obstetrician for termination of the pregnancy, but in view of her eagerness to have the child, and with consultation from other qualified medical specialists, the obstetrician accepted her as a pregnancy case. The predicted date of delivery was April 20, 1960.

In late November 1959, she was tested by a clinical psychologist. Her referral for this assessment was on a routine basis, the same as other pregnant women who were being seen as research subjects for the National Collaborative Project.

Shortly after the time of the psycological assessment, the patient developed a severe upper respiratory infection with considerable pain in her ears, high temperature, dizziness, and related symptoms. She was admitted to the Providence Lying-In Hospital for treatment of these ailments and was soon transferred to the Chapin Hospital for communicable diseases, diagnosed as a case of meningitis caused by pneumococcus bacteria. A ruptured abscess in the patient's middle ear sent septic emboli to the base of her brain, and she died very suddenly. At the time of her death, there was an audible fetal heart-beat, so a post-mortem Caesarean section was performed in the attempt to save the child's life. The infant girl lived for 18 hours, then died of hyaline membrane disease of the lungs, a condition peculiar to premature babies and the progeny of diabetics.

Having briefly reviewed this background and medical history, let us look at the psychological test findings in an attempt to discover the psychodynamics that were operating in this case approximately 2 months before her untimely death.

---

[1] This study was made possible by a research grant, No. B-2356, from the National Institute of Neurological Diseases and Blindness, Public Health Service, awarded to the Brown University Institute for Research in the Health Sciences. The present report stems from an ancillary study to the National Collaborative Project, directed locally by Glidden Brooks at Brown University, which is investigating perinatal factors in child development. We wish to express our indebtedness to Bertram H. Buxton, Jr., obstetrician on the Providence Lying-In Hospital staff, whose collaboration made this case study possible.

PROCEDURE

The findings to be reported were obtained on the basis of data gathered in a 2-hour, individual, testing session conducted by an experienced clinical psychologist. The patient was administered the following tests: (*a*) vocabulary and picture completion subtests from the Wechsler-Bellevue Intelligence Scale; (*b*) 50-item, true-false, manifest anxiety scale; (*c*) happiness self-rating scales; (*d*) 52-item sentence completion test, containing 26 first-person and 26 third-person stems; (*e*) Cards 1, 2, 3GF, 4, 7GF, 10, 12F, and 13MF from the standard Thematic Apperception Test (TAT); and (*f*) draw-a-person test. The three projective protocols were independently scored for reliability by two experienced clinical psychologists, and a highly satisfactory degree of agreement was found in the scoring and interpretation of this material.

At the completion of the testing session, on the basis of his clinical impressions, the psychologist rated the patient on 12 characteristics contained in a maternal personality rating scale. Each of the ratings was made on a 7-point scale, running from 1 at the socially undesirable end to 7 at the socially more healthy end. The 12 scores were summed, resulting in a total score indicative of a negative or a positive personality picture according to the clinical evaluation.

PSYCHOLOGICAL FINDINGS

On the basis of the intelligence test, it was found that the patient had an IQ of 140, indicating very superior intellect. This finding was supported by the psychologist's impression, based on the clinical interaction, that intellectually she was very bright and functioned at a high level.

On the manifest anxiety scale, the patient received a rather low score of 12. Results obtained from previous testing of normal, healthy, pregnant women have indicated a mean manifest anxiety score of about 16. Studies of randomly selected female, nonpregnant, college undergraduates have also reported mean scores of about 16 on this anxiety inventory. Thus, it appears that the patient was below average in the number of anxiety signs and symptoms she admitted experiencing.

The happiness self-ratings involved comparing oneself with most other people and indicating self-ratings along 4-point scales, running from "much less happy" to "much happier." In response to this procedure, the patient indicated that she regarded herself as "slightly happier" than most people. On the scale that required her to indicate the rating she believed she would receive from friends and family members, the patient again made the rating "slightly happier" than most people. In addition to these general happiness ratings, the patient was asked to make two specific ratings concerned with personal contentment with the state of pregnancy. In response to the question, "How happy or unhappy were you to learn that you were pregnant?" the patient indicated that she was "slightly happy." The second rating was made in response to the question, "Compared to the way you felt before you were pregnant, are you happier or unhappier at present?" and the patient rated herself as "slightly happier." Thus, on the basis of the patient's conscious avowal, it appears that she viewed herself as a fairly happy person, thought that she appeared happy to others, and was content to be pregnant.

Before considering the results obtained from the projective techniques, let us report the clinical psychologist's evaluation of the patient's personality structure according to his "maternal personality ratings." These ratings yielded a total score of 78, which is markedly higher than the mean rating of 50 that was obtained in previous studies of normal, healthy, pregnant women. The attributes on which the patient was rated consisted of such traits as contented, peaceful, cooperative, warm, self-assured, communicative, active, etc. This patient's high score indicates that she made a very favorable impression on the psychologist, and he regarded her as a very well adjusted, emotionally mature, and most adequate individual. In this regard, however, it should be pointed out that previous research has shown h ghly significant positive correlations between piatients' IQ, as measured by objective intelligence tests, and the favorableness of clinical evaluations as indicated by the maternal personality ratings.

Projectively, the patient's psychodynamics appear in a quite different light. The sentence completion test and the TAT were scored for the personality traits of egocentricity, pessimism, distrust, anxiety, and resentment, which previous research has shown to form a personality syndrome termed "alienation." Scoring the patient's sentence completions for indices of these personality dispositions resulted in an alienation syndrome score of 16, which is slightly higher than the mean of 13 obtained in previous analyses of responses obtained from normal pregnant women. More noteworthy, however, is the finding that 12 of the patient's alienation scores were based on completions she gave to sentences that were worded in the third person; only 4 of the alienation scores were in response to first-person sentence stems. It appears, therefore, that on a conscious level the patient presented a quite favorable personality picture, but in response to projective stimuli, she appeared considerably more "alienated" or maladjusted than most pregnant women.

A few examples drawn from her sentence completion test will serve to evidence the kinds of responses that revealed her conflicts and disturbances. In response to the stem, "Rita trusts——," she answered "nobody," which suggests a suspiciousness and distrust that were also indicated by other sentence completions. The stem, "Paula feels that the world is——," elicited the completion, "coming to an end," which is probably a sign of extreme pessimism. "I feel anxious when——" stimulated the response, "my sugar is low," and in response to the stimulus, "My greatest fear is——," she responded "getting sick." These projective responses, together with many others that were recorded, indicate that the patient was considerably more anxious, pessimistic, worried about her physical condition, and distrustful than she appeared on the surface.

The TAT revealed even greater evidence of psychopathology. The TAT stories were scored for indices of the five negative dispositions that form the "alienation syndrome." Scoring of this fantasy material resulted in an alienation score of 27, which is considerably higher than the mean alienation score of 18 obtained in previous studies of normal pregnant women. In addition to content, outcomes or story endings were also scored for positive or negative indices. The outcomes of four of this patient's five stories had negative outcomes. With the same weighted scoring system that was employed previously, she received a negative outcome score of 8, which is noticeably higher than our mean of 3 from earlier research. Thus, both the content of her fantasies, stimulated by the TAT pictures, and the nature of the endings she gave to her stories, indicate that this patient was noticeably more alienated, disturbed, depressed, and pessimistic than the majority of pregnant women studied previously.

Some excerpts from her TAT protocol will serve to further support the picture of emotional disturbance and inner turmoil that was gleaned from the projective material. In response to Card 2, the patient told a story about a family in which the husband was dead and the son and mother were left alone. She closed this story with, "The mother looks a little pregnant, but I won't put that in the story. She looks a little old to me." On an unconscious level, she apparently had a very dismal outlook on pregnancy and was trying to suppress any disturbing ideas she may have had about being pregnant. Since the patient was 39 years old, her statement that the pregnant woman looked a little old very likely had personal relevance. It is also noteworthy that this story, in which she tried to avoid coping with fantasies about pregnancy, contained a family in which there had been a death.[2]

The patient began her story to Card 3GF with the statement, "Oh my! This is a sad one." Then she told a most depressing story about a family in which the husband had left the wife alone with a young child. The child had been hurt in an accident, and the mother wanted to call the father but did not know where to locate him. The story ended with, "The doctor has said there is little hope that the boy will survive. It's a very depressing looking picture." Thus, here again, there are indices of family conflict, death, and depression.[3]

In telling a story about the female figure in Card 12F, the patient said that, "Outwardly, she appeared to be very kind and proper, but actually she had deep seated hatreds for most people, and if you could see inside her soul, she would look like the old hag that she really was inside." It seems, then, that this patient was concerned about differences between surface appearance and unacceptable motives operating at deeper levels. The reference to the person's "insides" seems particularly noteworthy in view of her own history of surgery, which undoubtedly caused much more emotional difficulty than appeared on the surface. Moreover, this TAT fantasy provides signs of hostility and resentment that were not evident in the patient's responses to the more objective assessment procedures.

In response to Card 13MF, which usually stimulates sexual fantasies but also is apt to evoke stories of aggression from some individuals, the patient constructed a story in which love, sex, hate, and death were intermingled. The story was about a man who was passionately in love with a very beautiful girl. The girl loved him very much, but he was insanely jealous of her and suspected her of being insincere and untrue to him. In a fit of rage, he killed her. The story was ended on the tragic note that, "He was hit by a great feeling of remorse because he knew he was sorry, that she really did love him; but it was too late then." Again we see an unsatisfying relationship between man and woman, with distrust, extreme hostility, death, and depression. These examples from the TAT should suffice to indicate the overwhelming evidence of alienation and psychopathology revealed by the patient's fantasies. The content was consistently morbid and conflict laden, with uniformly desperate outcomes.

---

[2] In looking into the patient's background, a noteworthy anecdote was uncovered. Shortly before the patient was seen for psychological testing, in November, two of her neighbors had died. One was a 38-year-old male who died from ulcers; the other was a 70-year-old woman who died from natural causes. It has been reported that about that time the patient told a close friend that death usually comes in threes, and she believed that she was destined to complete the trio.

[3] After the present report was prepared, it was discovered that this patient's husband had been divorced from his first wife, and, moreover, that he suffered from a manic-depressive condition requiring psychiatric treatment.

The final projective material to be considered here was obtained from the draw-a-person test. The patient drew a picture of a girl with the corners of the mouth turned up in a sort of forced smile and with both eyes tightly closed. A plausible clinical interpretation of this drawing is that the figure was trying to shut out forcefully all awareness of what was transpiring both within and around her and was straining to provide a happy outer appearance. The drawing also contained evidence of considerable anxiety, indicated by a great amount of shading, many irregular lines throughout the figure, and much retracing of lines in the contour of the figure. Thus, both on the basis of objective signs and clinical interpretation, the figure drawing suggests that the patient relied very heavily on defense mechanisms of denial and repression to control her severe inner turmoil and conflict.

From the data obtained from this comprehensive battery of direct and projective assessment techniques, it now seems quite evident that in this patient there was great discrepancy between surface and depth. On the basis of her intelligence test performance, her conscious avowal of manifest anxiety symptoms, happiness self-rating, and verbal interaction with the psychologist, she appeared to be functioning very adequately. However, the projective techniques revealed that beneath this seemingly healthy exterior there was a great store of anxiety, distrust, resentment, and pessimism. In this case, therefore, it is essential to compare the findings derived from both the direct and the projective methods in order to acquire a valid picture of her personality dynamics.

In this regard, one wonders if the picture of emotional maturity, confidence, intelligence, and mature control which she presented, even to an experienced clinical observer, may have served to influence the obstetrician and helped him to accept this physically sick woman as a good risk for pregnancy and normal childbirth. It seems possible that if she had appeared anxious, worried, and doubtful in their interpersonal relationship, in view of her physical condition and the risk involved, the obstetrician might have been more reluctant.

It seems warranted to conclude that this patient was considerably more worried and unhappy than she appeared, and that it was a rather exhausting struggle to maintain her emotional control. Just what having this baby meant to her is not certain. The projective tests provide strong suggestions that her marital relationship may not have been a happy one, and she may have viewed the baby as a means of holding the family together. It is also possible that she wanted to have the baby for her own sake, as a result of some form of counterphobic mechanism, denying fate and reality. Giving birth, at this stage of her existence, would serve to prove that she was still able to function as a woman, that she was not merely existing in the shadow of death but was beginning a new life. At any rate, she wanted to have the baby and fought to minimize the emotional upset generated by the serious risk that was involved.

It may be that this tragic case will make some form of contribution to medical science. Our purpose, as psychologists, however, is to show the sort of contribution that intensive psychological assessment can make in medical research settings. Findings from unique cases such as this may help us toward our goal of ultimately attaining greater scientific understanding of human psychodynamics. From a practical viewpoint, this case vividly emphasizes the need to include both direct (surface) methods and projective (depth) methods in any thorough personality assessment

SUMMARY

This paper presents a psychological case study of a 39-year-old woman who died during pregnancy. Previous to the pregnancy, she had undergone major surgery for stomach cancer and had been strongly advised not to become pregnant. During the fourth month of pregnancy she was administered a comprehensive battery of psychological tests. This testing was done in connection with a National Collaborative Project on perinatal factors in child development, and at the time of testing she was regarded as merely another case in a large sample of pregnant women who were to be included in the research program. During the sixth month of pregnancy, however, the patient died from a brain abscess associated with meningitis, and the baby succumbed following a post-mortem Caesarian delivery. In view of the uniqueness of the case, we conducted a detailed examination of the psychological test findings in the attempt to learn about psychodynamic factors that may have been operating shortly before her death.

The patient's performance on the Wechsler-Bellevue Intelligence Scale showed her to have an IQ of 140. On the manifest anxiety scale, her score indicated fewer anxiety symptoms than the mean number reported in previous groups of normal pregnant women. On a series of self-rating scales,

the patient indicated that she was happier than most people, believed that other people also viewed her as happier than most people, and that she was happy to be pregnant. The examining psychologist's ratings of the patient, on a series of maternal personality characteristics, resulted in a score markedly higher than the average score obtained in a previously studied sample of pregnant women, indicating a very positive personality picture according to the clinical evaluation.

On the contrary, however, findings derived from the TAT, sentence completion test, and draw-a-person test, provided abundant evidence of psychopathology underlying the surface picture of healthy emotional adjustment. It was found that the patient was extremely anxious, distrustful, resentful, and pessimistic. Thus, the projective material revealed that she was an alienated and conflict-laden individual who made strong use of the mechanisms of denial and repression in the attempt to hide her inner turmoil both from her own conscious awareness and from the awareness of others. It was concluded that these case findings serve to reaffirm the importance of including both direct methods and projective methods in any thorough assessment of personality dynamics.

(Early publication received March 21, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 292-294

CRITIQUE AND NOTES

THE COMPLEMENTARITY OF PERSONALITY NEEDS IN FRIENDSHIP CHOICE

MARY ST. ANNE REILLY, WILLIAM D. COMMINS, AND EDWARD C. STEFIC
Catholic University of America

EARLY studies in the dynamics of friendship agreed that friendship was characterized by a positive, though slight, resemblance between friends. However, the personality theories of Murray (1938) and Rogers (1951) suggested that friendship involves a complementary relationship of mutual need satisfaction, and Winch, Ktsanes, and Ktsanes (1955) found evidence of complementary needs in mate selection. The present study, concerned with investigating the complementarity of personality needs in friends of the same sex, proposed these questions: Are the need patterns of friends complementary? Are the personality needs of friends mutually satisfying?

The two main hypotheses proposed by Winch et al. (1955) served as a framework. These were that (*a*) a person high in one need would be attracted to someone low in that need (complementarity of needs), and (*b*) a person high in one need would be attracted to someone high in the opposite need (need satisfaction). In addition to investigating the actual, self-perceived need patterns of friends, we were interested also in studying the phenomenal need patterns of friends—their needs as perceived by each other. Here, we were following Rogers' (1951) suggestion: "it is the perception, not the reality, which is crucial in determining behavior." Finally, to establish a basis for comparison with earlier studies in the field, of which some have found friends to be similar in values or evaluative attitudes, we proposed a third hypothesis, that friends would tend to be similar in values.

METHOD

Two groups of resident female Catholic college students volunteered as mutual friend pairs, with 25 junior pairs and 25 sophomore pairs (total $N = 100$). A control group, set up by randomly matching the juniors with the sophomores by a table of random numbers, formed a contrasting group of nonfriends. The age range of the subjects (*S*s) was from 18 years 11 months to 21 years 1 month, with a mean age of 19 years 11 months.

The first instrument used was the Edwards Personal Preference Schedule (PPS), which provides measures of 15 personality needs. The PPS was designed to minimize the factor of social desirability, by employing a forced-choice format. The second instrument was the Allport-Vernon Study of Values (1931), which provides relative measures of six basic interests or values. (The 1931 edition, instead of the revised edition was used, as a better basis of comparison with earlier studies in the area.)

The PPS was administered to all junior *S*s, to obtain self-perceived need scores (each *S* took the test as she perceived herself). During the same testing period, the Allport-Vernon was also administered. Two days later, the same *S*s took the PPS, but this time as each thought her friend would answer. This yielded friend-predicted need scores for all *S*s. Two weeks later, all sophomore *S*s took the tests, following the same procedure.

From the raw scores of the personality needs measured by the PPS, Pearson product-moment correlations were computed for the following two variables of friends: (*a*) the self-perceived needs of friends, and (*b*) the phenomenal needs of friends (an *S*'s self-perceived needs with that *S*'s prediction of her friend's needs); and one variable on nonfriends: their self-perceived needs. From the raw scores on the six Allport-Vernon values, correlations were computed for an additional variable on friends: self-perceived values.

RESULTS AND DISCUSSION

The first hypothesis, about complementary needs between friends, was tested via the following questions:

> Does a person have for a friend, one whose actual, self-perceived needs complement her own?
>
> Is there any difference between the correlation of a friend pair's actual self-perceived needs, and that friend pair's phenomenal needs (the correlation of a person's actual needs with her prediction of her friend's needs)?
>
> What is the relationship of the personality needs of nonfriends?

The following correlations on friends and nonfriends were inspected: complete need profiles, each personality need with the same need in the partner, and the highest need of each *S* with the same need in the partner. A complementary relationship between the personality needs should be indicated by negative correlations.

Need Profiles

In respect to the 50 correlations of the actual, self-perceived need profiles of the 50 friend pairs, and those of the 50 nonfriend pairs, the number of negative *r*'s was the same for both groups. The range of the *r*'s for friends was from −.54 to .85

with a mean r of .23, while the range for nonfriends was from −.54 to .85 with a mean r of .18, the difference obviously falling short of significance. This finding was not similar to that of Winch et al. (1955), who reported a significant difference between the mean r of .10 on husband-wife pairs in a need interview, and .25 for randomly matched men and women. Admittedly, the husband-wife relationship may be more of a complementary one than that between friends.

An inspection of the 100 correlations of friend pairs' phenomenal need profiles shows that the proportion of negative correlations here was almost the same as for the self-perceived need profiles of friends; the range of r's was from −.73 to .90 with a mean r of .25. This value of course does not differ significantly from that of .23 for the self-perceived need profiles of friends.

From the data, then, it is clear that the self-perceived need profiles of friends were not complementary to one another. There was no significant difference between friend pairs' actual needs and the phenomenal needs of friend pairs. Need profiles of nonfriends did not differ significantly from those of friends.

If these data in regard to need profiles did not support the first hypothesis of a complementary relationship between friends, neither did they warrant any assertion as to the similarity of friends in respect to personality needs. In the first place, though the mean r of the correlations of friends' need profiles was positive (.23) it was not significantly different from that of nonfriends (.18).

TABLE 1

CORRELATIONS OF FIFTEEN PERSONALITY NEEDS

Personality Needs	Friends' Actual	Friends' Phenomenal	Nonfriends' Actual
Achievement	.09	.24*	−.07
Deference	.16	.25*	−.13
Order	−.05	−.03	.01
Exhibition	.06	.22*	−.14
Autonomy	.11	.31**	.13
Affiliation	.16	.20*	.18
Intraception	−.07	.14	.04
Succorance	.16	.19*	−.10
Dominance	.17	.15	.11
Abasement	.13	.35**	.05
Nurturance	.02	.30**	−.02
Change	−.15	.10	.12
Endurance	−.15	−.09	.15
Heterosexuality	−.17	.32**	.09
Aggression	.08	.29**	−.14

* Significant at the .05 level.
** Significant at the .005 level.

TABLE 2

CORRELATIONS OF OPPOSITE NEEDS OF FRIEND PAIRS

Opposite Needs	Actual	Phenomenal	Opposite Needs	Actual	Phenomenal
Ach.-Def.	.02	−.12	Aut.-Suc.	.01	−.09
Ach.-Aff.	.02	.15	Aut.-Aba.	−.07	−.09
Ach.-Suc.	.00	.10	Aut.-Nur.	−.01	−.24
Ach.-Aba.	.07	−.11	Aff.-Dom.	.05	−.05
Ach.-Nur.	−.01	.03	Aff.-Agg.	.11	−.12
Def.-Exh.	−.17	.07	Dom.-Suc.	−.15	−.19
Def.-Aut.	−.13	−.03	Dom.-Aba.	−.09	.16
Def.-Dom.	.08	.00	Dom.-Nur.	.02	.52
Def.-Agg.	.01	−.12	Agg.-Aba.	.02	−.15
Aut.-Aff.	.06	−.29	Agg.-Nur.	.12	−.08

Separate Personality Needs

The second step in the analysis sought to discover whether friends tend to be more complementary to one another in regard to certain personality needs than to others. To this end, correlations for each of the 15 personality needs from the data on the self-perceived needs of friends and of nonfriends, as well as from the phenomenal needs of friends, were inspected (see Table 1). The double entry method of computing product-moment correlations for each of the 15 personality needs was used, as recommended by Goodenough and Anderson (1931), for measuring an interchangeable variable. No consistent complementary relationship between friends on personality needs was indicated. In other words, though some friend pairs were complementary on need endurance, for example, other pairs were similar in regard to this need. There was no significant difference between the r's of friends and nonfriends; or between the r's of friends' self-perceived needs and their phenomenal needs.

Each S's Highest Need

The data were next analyzed to find out whether the need on which the S obtained the highest score was complementary to the same need in the partner. S's highest need correlated .02 with the same need from the friend's self-perceived needs, .01 with the prediction of her friend's need, and −.01 with the nonfriend's self-perceived need, values not significantly different from one another or from zero. There was no evidence, therefore, of a complementary relationship between a person's highest need and that same need in her friend.

To test the second hypothesis about mutual need satisfaction between friends, opposite needs of friend pairs were compared, our prediction being that they would correlate positively. For this problem, a list of 20 sets of opposite needs

TABLE 3
VALUES OF FRIEND PAIRS

Values	*r*	*M*	*SD*
Theoretical	.01	24.8	6.18
Economic	.09	23.4	6.92
Aesthetic	.14	27.3	7.16
Social	.06	29.4	5.01
Political	.15	30.5	5.92
Religious	.03	42.6	6.82

was devised from statements from Murray (1938, pp. 151–181). (Edwards has reported in the table of intercorrelations in his PPS manual, all negative *r*'s, ranging from −.12 to −.36, on the opposite needs which were used in this study. However, these *r*'s are not comparable to those presented in the present study, since Edwards' *r*'s are intra-individual, while the *r*'s obtained in this sample are between friends.) The data were analyzed in regard to the self-perceived needs of friends, and the phenomenal needs of friends, as is indicated in Table 2. Though a number of positive *r*'s were obtained, these were too low to support the hypothesis of mutual need satisfaction between friends. Here again, this finding was not what might have been expected as a result of the findings of the somewhat similar study on mate selection of Winch et al. (1955). They reported, for example, that "men high in need achievement tend to select as wives, women who are high in need abasement, need deference, need succorance, and need nurturance." We suggest that while the essential nature of the heterosexual relationship of husbands and wives may be a complementary one, that of friends of like sex is not essentially complementary.

In this sample, data on the values of friends from the Allport-Vernon Study of Values were very similar, on the whole, to earlier studies in this area. In regard to Richardson's study (1940), for example, the *r*'s were similar, except for theoretical and religious values, for which he found *r*'s of .21 and .34, respectively. However, because the present sample was more homogeneous, restriction of range could account for the lower correlations. The means of each of the values were also similar in the two studies, the only exception being that for religious value, which was 42.6, significantly higher than the 31.4 reported by Richardson. The data indicate a positive though slight resemblance between friends in values, as can be seen in Table 3.

SUMMARY AND CONCLUSIONS

This study sought to find out whether friendship involves a complementarity of personality needs, and whether these needs are mutually satisfying. Need profiles from the Edwards PPS were obtained for each of 50 friend pairs. Friend-predicted need profiles from the PPS were also obtained. Relative scores of six values from the Allport-Vernon Study of Values were secured for each *S*. Correlations were computed for complete need profiles, for single personality needs, for opposite needs, for the three following groups: friends' self-perceived scores, friends' self-perceived with friend-predicted scores, and randomly matched pairs' self-perceived scores.

The following conclusions were drawn:

1. In this sample, no consistent complementary relationship was found in regard to self-perceived personality needs of friends, nor was there any evidence of mutual need satisfaction between friends.

2. Friends, in general, did not tend to see themselves and their friends as more consistently complementary rather than similar.

3. On the other hand, there was no conclusive evidence for a relationship of similarity of personality needs of friends.

4. Friends tended to be slightly similar in values.

REFERENCES

GOODENOUGH, FLORENCE L., & ANDERSON, J. E. *Experimental child study*. New York: Century, 1931.

MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

RICHARDSON, H. M. Community of values as a factor in friendship of college and adult women. *J. soc. Psychol.*, 1940, **11**, 303–312.

ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.

WINCH, R. F., KTSANES, T., & KTSANES, VIRGINIA. Empirical elaboration of the theory of complementary needs in mate selection. *J. abnorm. soc. Psychol.*, 1955, **51**, 508–513.

(Received June 22, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 295–297

CONTEXT EFFECTS IN SOCIAL PERCEPTION[1]

LEON H. LEVY
Indiana University

BEGINNING with the now classical study by Murray (1933) on the effects of fear arousal on judgments of maliciousness in photographs, a fairly extensive body of literature has developed concerned with the personal determinants of social perception. In contrast, little is known concerning the situational determinants of social perception, especially those which do not seem to have a readily apparent effect upon one of the presumed personal determinants. That is, excluding situational factors that might be expected to induce some affective response in the perceiver such as fear, frustration, or hostility, and in contrast to our knowledge of the effects of various field factors such as background illumination and ambient noise level upon perception of relatively simple visual and auditory stimuli, our knowledge of the pertinent situational determinants of social perception is grossly deficient. Yet there must be such determinants and any theory of social perception cannot be considered complete without taking them into account.

Hunt (1941), Helson (1948), and others (Rudin & Stagner, 1958) have argued that laws of judgment exist which are not unique to the type of stimuli judged. Hunt (1941), attempting to establish this point, demonstrated shifts, allegedly reflecting anchor effects, in judgments using an 11-point rating scale in which the ends were either anchored or not depending upon the experimenter's (*E*'s) instructions to the subject (*S*) regarding their meaning. Unfortunately, the results obtained by this procedure would seem to be just as easily accounted for by alterations in verbal response probabilities brought about by the instructions, as by alterations in perceptual processes.

The present study continues within the vein of that by Hunt (1941) but differs from his in that it studied anchor effects where these were effected entirely through modifications of the stimulus context without alterations in verbal instructions. Finding predictable judgmental shifts with this type of procedure would seem to present an especially strong argument for the presence of field determinants of social perception as well as for the nonspecificity of the laws of judgment. This approach clearly places the locus of effect within the stimulus situation, obviates the problem of dealing conceptually with verbal instruction, and reduces the applicability of an interpretation in terms of response probability.

The hypothesis to be tested in this study is that judgments of a target photograph vary inversely with the context of other photographs with which it is presented, such that the greater the probability that the context will elicit judgments in one direction on a particular bipolar dimension, the greater the probability that the target will be judged in the opposite direction. This hypothesis is predicated upon the assumptions that (*a*) judgments of this type are a function of the location of a point of subjective equality or adaptation level (Helson, 1948) within any given dimension and (*b*) that this point varies directly with the value of the stimuli comprising the judgmental context.

The hypothesis was tested by using four bipolar dimensions, good–bad, dangerous–safe, reliable–unreliable, and tense–relaxed, and three contexts for each designated as high, medium, or low in terms of the estimated probability that the stimuli comprising the context would be judged as good, dangerous, reliable, or tense, respectively.

METHOD

Stimuli. The photographs used in this study were of white, male applicants for admission to graduate training in psychology at Indiana University who were either not accepted or did not elect to come. Each of these pictures had been used in a previous study (Levy & Dugan, 1960) in which 100 *S*s had judged them in terms of the traits used in this study as well as other traits. Since, in this previous study, these pictures were part of a total of 225 presented in random order and judged on one of 15 bipolar dimensions, it did not seem likely that context effects, as they were to be studied here, entered to any great extent in the judgments obtained. Thus records of the frequency of judgments made for each picture in the previous study served as the basis for selecting pictures that comprised the various contexts and targets in the present study.

For each dimension, seven pictures were used, two for each context and one as the target, with no picture being used in connection with more than one dimension. Each picture was mounted on a 3 × 4 inch piece of poster board with a label containing the two terms of the dimension attached below it. Judgments of photographs constituting the high context condition had a frequency range between 80% and 98% in the specified direction for the dimension in question, for medium con-

---

[1] This investigation was supported in part by a Research Grant (M-2067) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

The author wishes to express his gratitude to Jane A. Kelly who ran the subjects of this study.

text pictures the corresponding judgments ranged between 46% and 56%, and for low context pictures between 10% and 35%. Target photographs ranged between 49% and 52% in judgments placing them at one particular end of their respective dimensions.

Subjects. The *S*s used in this experiment were 120 students, 78 men and 42 women, drawn from an introductory psychology course. They were assigned at random to one of four groups of 30 *S*s each, with the proportion of men and women approximately the same between groups. None of these students had participated in the previous study using the photographs.

Procedure. Each *S* was seen individually and presented with three sets of three pictures each, one set at a time. All pictures within a set were judged on the same dimension and consisted of two context pictures and one target. Each set represented a different dimension as well as a different context level, and within each group the order of presentation of context levels, and consequently dimensions, was systematically counterbalanced. Thus all *S*s in any given group made judgments in terms of only three of the four dimensions and in connection with only one corresponding context for each, but over all four groups, each dimension was involved at each context level equally often yielding 30 judgments for each target at each context level of its respective dimension.

On each presentation, the set of three pictures was presented on a small stand directly in front of the *S* with the target picture flanked by each of the context pictures. *S*s were instructed to look at all three pictures and then to give their first impressions of each in terms of one of the two adjectives on the label below the picture. As soon as this was done and recorded by *E*, the next set was presented. From the *S*s' standpoint, the target pictures were in no spurious way distinguishable from their contexts.

RESULTS

The number of *S*s judging each target as good, dangerous, reliable, or tense, was tabulated for each context level. These data, presented in Table 1, reveal the monotonic increase predicted from the hypothesis in all instances except the good–bad dimension where a reversal between the high and medium context levels occurred. A Friedman two-way analysis of variance by ranks (Siegel, 1956) applied to these data in order to determine whether judgments differed significantly between context levels over all four dimensions, yielded a χ^2 of 6.5 which is significant at the .04 level. Further, testing for the significance of the difference in proportions of judgments in a specified direction between the high and low context levels for each dimension separately indicated a difference significant at the .04 level for the dangerous-safe dimension and at the .003 level for the tense-relaxed dimension, using a one-tailed criterion in each case since the direction had been predicted. Thus, these results clearly support the hypothesized inverse relationship between the probability of particular judgments of the context photographs and of targets contained in the context.

TABLE 1

FREQUENCIES OF SUBJECTS MAKING SPECIFIED JUDGMENTS OF TARGET PHOTOGRAPHS IN VARIOUS CONTEXTS

Dimension	Context Levels			Significance of proportion of high vs. low judgments
	High	Medium	Low	
Good–Bad	13	12	16	*ns*
Dangerous–Safe	14	17	20	.04
Reliable–Unreliable	12	15	18	*ns*
Tense–Relaxed	8	16	18	.003

Note.—The frequencies given are for *S*s giving the term in italics as their judgment. In each case the total possible judgments is 30.

DISCUSSION

In the language of social psychology, the present findings have most immediate relevance for the concept of reference groups (Hyman, 1942; Kelley, 1952; Merton & Kitt, 1950; Newcomb, 1950; Sherif, 1948). For the *S*s of this experiment, the context pictures may be seen as functioning as reference groups for their judgments of the target pictures. From this standpoint, the results clearly point to the potency of the reference group as a standard of comparison in social perceptual situations even where the likelihood is minimal that the person will encounter any sanctions for ignoring it. Further, membership in a group is clearly not a necessary condition for its functioning as a standard of comparison for the individual. Indeed, the principles governing the operation of reference groups in social perception appear to be no different from those operating in classical psychophysical situations involving judgments of standard and comparison stimuli.

Certain questions that might be raised in connection with the operation of reference groups would seem to be ideally suited to investigation by the method used in the present study. By appropriate manipulations of the constitution of the set of context pictures, the effects on judgments of such variables as the size of the context, context-to-target social relationships of various kinds, heterogeneity of contexts, etc. are open to investigation. The conditions under which context effects are maximal and minimal may also be studied by this method.

Implicit in the latter problem is the question of individual differences in susceptibility to context effects and their personality correlates. The extent to which the context or reference group affects a person's judgment would seem akin to the question of field dependency in personality theory (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954), and thus an opportunity exists, by means of the present technique, to test the generality of this construct and also to effect a further *rapprochement* between social psychology and personality theory.

Personality theories in which the self-concept plays a central role may also find the results of the present study of some relevance. For if context affects judgments of others, it is reasonable to believe that it would have a similar effect upon self-perception. Efforts to account for a person's self-concept at any given time, as well as attempts at altering it, might therefore do well to consider the context within which the person judges himself as well as his personality dynamics and life history.

SUMMARY

Concerned with the problem of field determinants of social perception, this study investigated contextual effects on judgments of photographs of persons. Photographs, designated as targets, were presented in contexts of two other photographs which were expected, on the basis of previous research, to vary in their probability of eliciting a particular judgment along a specified bipolar trait dimension. Data obtained clearly supported the hypothesized inverse relationship between the probability of the context eliciting a particular judgment and that of the target eliciting a similar judgment. The results were discussed in connection with reference group theory in social psychology and the constructs of field dependency and the self-concept in personality theory.

REFERENCES

HELSON, H. Adaptation level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

HUNT, W. A. Anchoring effects in judgment. *Amer. J. Psychol.*, 1941, **54**, 395–403.

HYMAN, H. H. The psychology of status. *Arch. Psychol., NY*, 1942, No. 269.

KELLEY, H. H. Two functions of reference groups. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (Rev. ed.) New York: Holt, 1952. Pp. 410–414.

LEVY, L. H., & DUGAN, R. D. A constant error approach to the study of dimensions of social perception. *J. abnorm. soc. Psychol.*, 1960, **61**, 21–24.

MERTON, R. K., & KITT, ALICE S. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Studies in the scope and method of "the American soldier."* Glencoe, Illinois: Free Press, 1950. Pp. 40–105.

MURRAY, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, **4**, 310–329.

NEWCOMB, T. M. *Social psychology.* New York: Dryden, 1950.

RUDIN, S. A., & STAGNER, R. Figure-ground phenomena in the perception of physical and social stimuli. *J. Psychol.*, 1958, **45**, 213–225.

SHERIF, M. *An outline of social psychology.* New York: Harper, 1948.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

(Received July 2, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 297-300

THE FORMATION OF POSITIVE ATTITUDES TOWARD GROUP MEMBERS[1]

BERNICE EISMAN LOTT
Kentucky State College

AND

ALBERT J. LOTT
University of Kentucky

In an attempt to predict group behavior on the basis of general psychological principles as contrasted with formulations specific to the area, the concept of group cohesiveness has been re-examined and reformulated within a learning theory framework (Lott, in press). Within this framework, cohesiveness is defined as *that group property which is inferred from the number and strength of mutual positive attitudes among the members of a group.* The concept of attitude is used, instead of the more usual one of attraction, because of its precise and particular meaning within learning theory, i.e., an implicit anticipatory response having cue and drive properties (Doob, 1947). By defining cohesiveness in terms of mutual positive attitudes among group members, the assumption is made that the members who comprise a group constitute its most significant components.

A number of hypotheses regarding both the antecedents and consequents of cohesiveness, as defined above, have been derived from learning theory (Lott, in press). The problem of this experiment has been to test the most fundamental of these hypotheses, one which concerns the conditions under which positive attitudes toward group members may be formed.

---

[1] A paper, based on this investigation, was read at the 1959 American Psychological Association meetings in Cincinnati.

It is predicted that if a person is rewarded in the presence of others (fellow group members), he will develop positive attitudes toward them. This proposition rests upon the following assumptions:

1. Persons may be conceptualized as discriminable stimuli to which responses may be learned.

2. A person who experiences reinforcement or reward for some behavior will react to the reward, i.e., will perform some observable or covert goal response (R_g or r_g).

3. This response to reward will become conditioned, like any other response, to all discriminable stimuli present at the time of reinforcement.

4. A person (group member) who is present at the time that Individual X, for example, is rewarded thus becomes able, in a later situation, to evoke R_g or, what is more likely, its fractional and anticipatory component, $r_g - s_g$. This latter response, which Hull has called "expectative" (1952, Ch. 5), was earlier interpreted by Doob (1947) as the underlying mechanism of an attitude.

The specific prediction tested by this study is that members of three-person groups who are rewarded for their performance in the presence of their fellow group members will more likely develop positive attitudes toward them than will members of such groups who are not rewarded. Positive attitudes are inferred, here, from choices made on a sociometric test subsequent to, and outside of, the experimental situation.

METHOD

Subjects

Forty-eight children from the University of Kentucky Elementary School, 24 each from Grades 3 and 5, served as *S*s.[2]

Procedure

The *S*s were divided into 16 three-member groups, following the administration of two sociometric tests. These tests were given, by the regular classroom teachers, on two consecutive days, several days before the actual experimental situation. On the basis of the test results the groups were formed so that each group was made up of children who had *not chosen each other* on either of the tests. Four all male and four all female groups were formed from Grade 3; five male and three female groups were formed from Grade 5.

Omitting the preliminary instructions, the criterion question asked in the first test was as follows:

> *Test I*. Let us suppose that each one of you gets picked to take a trip to the moon in a rocket ship. This is a very special trip .... It is important for everyone who is on the same ship to get along well with each other .... Because of this you get a chance to pick two children to go along with you. Now, of all the children in this class, which two would you pick to travel in the same rocket ship with you?
>
> *Test II*. [For this test the children were asked to think of themselves as visitors from Mars and to choose two classmates whom they would like to have waiting for them when their spaceship landed on Earth.]

For a group situation in which rewards and nonrewards could be manipulated, a board game called "Rocket Ship" was devised. The object of the game, played by groups of three, is to land cardboard rocket ships on planetary objectives. Each objective is reached by traversing a separate path containing four danger zones, at each of which a choice between a white and a striped subpath (one "safe" and the other "dangerous") must be made. By having the children in a group take turns crossing the danger zones first, the *E*s could arrange to have some children succeed and others fail in reaching the planets safely. The manner in which this was accomplished will be described below.

On the day of the experiment proper (separate days for the two classes), the *E*s were introduced to *S*s as having developed a children's game which they wished to test. One group at a time was called upon (in a predetermined random order) to play the game in a room adjoining the regular classroom.[3] The following instructions, given to each group by one of the investigators (E_1), describe the manner of playing:

> ... You three are rocket ship pilots. Each one of you has your own ship but you are going to take a trip into outer space together, side by side. The first trip you are going to take is to ——. In order to get there you must follow this path which scientists have decided is the safest way to go. The scientists also know that, at a few points along the way on this path, there is great danger. When you get to one of these dangerous points you are going to have to stop and decide whether the striped path or the white path is the dangerous one. One of the paths will get you past the danger safely. If you take the wrong path your rocket ship will be blown up and you'll have to parachute to Earth. ... When you get to a danger point you'll have to decide which one of the group will take a chance and be the first to try either the striped or white path. If he gets through safely, the other two rockets may follow him. If his ship should get blown up, though, the other two ships will take the other path which you will know to be safe. You'll know that a ship has been blown up when

---

[2] The authors wish to acknowledge the cooperation of the following persons whose kind assistance made possible both the reported experiment and the pilot work which preceded it: M. Hitch, Principal of Rosenwald Laboratory School, Frankfort, Kentucky; A. Wolfe, intermediate grade teacher at Rosenwald; E. Sasman, Principal of University Elementary School, Lexington, Kentucky; A. Boone, M. Moore, and O. Barrett, teachers at University School of the third, fourth, and fifth grades, respectively.

[3] Wherever a child, preselected for a particular group, was absent on the day of the experiment, a substitute who met the criterion of nonchoice by, and of, the other group members, was, in all cases, available.

> you hear the sound of a bell, like this [sounded by E_2]. . . . Remember, you must take turns being first.
> . . . It's possible for all three of you to get through to a planet safely, but it may be that only two of you will make it, and maybe just one or none of you will make it. . . . If you reach a planet safely you will be able to choose one of these prizes [small plastic auto models] which you may keep. . . . We'll play half the game this morning and half this afternoon.

Each group tried for three objectives in the morning and three in the afternoon. Half the total number of *S*s was permitted to land safely on four planets (two at each session) while half was prevented from reaching any of them. "Reward" was thus defined as the receipt of four plastic car models (one for each successful landing). A child was either "rewarded" or "not-rewarded," i.e., made no successful landings at all and received no prizes; there were no in-between conditions. Prizes won in the morning session were held by the *E*s until the end of the afternoon session at which time the rewarded *S*s returned to their classroom with four model cars and the nonrewarded *S*s returned with none.

Rewarded and nonrewarded *S*s were selected on a random basis prior to the game. A nonrewarded *S* always had his ship "blown up" when, at his chance to go first at a danger point, he took either the striped or white path.[4] As soon as he made his choice of path, the "blow up" bell was rung by E_2 who sat somewhat apart from the game area.

Each group had been randomly assigned, prior to the game situation, to one of the following conditions which describe the number of group members who were to be rewarded during the game: zero, one, two, all. Four groups, two from each class, were assigned to each condition. This aspect of the design was introduced to avoid having the *S*s suspect that the game was "rigged." The natural flavor of the game was maintained by having *S*s know that it was possible for all three group members to be successful, or only two, etc.

Shortly before the close of the school day, approximately one hour after the last group had played the game, the classroom teacher administered another sociometric test (III), as follows:

> Suppose your family suddenly got the chance to spend your next vacation on a nearby star out in space . . . you can invite two children to go on the trip . . . and spend the holiday with you on the star. Which two children in this class would you choose to take with you?

After the choices were collected by the teacher, E_2 appeared before the class to thank the *S*s for their cooperation and help. And, because everyone had been such "good sports," four prizes were distributed to each of the youngsters who had not won them during the game.

TABLE 1

CHOICES MADE BY SUBJECTS ON SOCIOMETRIC TEST III

Subjects	Choices: Play-group member	Choices: Non-play-group member	N^a
Rewarded	11	37	48
Nonrewarded	3	45	48

[a] N = number of choices made; each *S* made two choices.

RESULTS

The results of the final sociometric test (III), which succeeded the play-group experience, are presented in Table 1. The proportion of play-group members chosen by rewarded *S*s was found to be significantly greater than the proportion chosen by nonrewarded *S*s. The obtained critical ratio, corrected for continuity, is 2.14 ($p = .03$; two-tailed). This confirms the prediction that *S*s who had been rewarded would choose members of their groups, on the final sociometric test, significantly more often than *S*s who had not been rewarded.

DISCUSSION

The present findings indicate that the formation of positive attitudes toward persons is predictable from learning theory principles and can be studied in the laboratory. This study thus extends the applicability of a general S-R framework to significant social behavior.

The prediction that positive attitudes toward persons can be formed by experiencing reward in their presence, was clearly confirmed. Since it is in terms of such positive attitudes among the members of a group that cohesiveness has been defined, the present experiment is seen to be concerned with the antecedents or determinants of cohesiveness even though no attempt was made to measure the variable directly. The specific concern of this first study was with the development of positive attitudes toward group members and not with the group property that results from such attitudes when they are mutual.[5]

There have been comparatively few experimental attempts to vary group cohesiveness other than by suggesting to the members of a group that they will like each other. One investigation in which determinants were experimentally manipulated (Thibaut, 1950) found a positive relationship between cohesiveness and group status. In another study, more relevant to the present one

---

[4] Each child was in the position of "being first" at a danger point at least once on his way to each of the planets since *S*s had been instructed to take turns and there were four danger zones per path and three children in a group. It was always possible, therefore, to "blow up" nonrewarded *S*s without interfering with a group's spontaneous behavior regarding which child would go first at which point.

[5] An operational measure of cohesiveness which follows directly from the definition of the concept given earlier has been developed by the investigators and will be reported in connection with its use in another study.

(Bovard, 1951), a significantly greater level of interpersonal affect was found in group-centered as compared with traditional leader-centered classes. Bovard suggested that this result was due to the fostering, in group-centered classes, of member-to-member interaction which produces greater accuracy in intermember perception, creating, according to Bovard, a situation conducive to "need satisfaction." That the experience of need satisfaction (reward) in the presence of group members can, indeed, result in positive affect toward those group members has been demonstrated in the present experiment.

The results obtained in this study should be evaluated in the light of the following factors which could only have tended to work against substantiation of the hypothesis:

1. Play-group members were, by design, neutral or negative stimuli for one another at the beginning of the experimental situation. This was assured by placing, in a group, only *S*s who had not chosen one another on two previous sociometric tests.

2. The amount of reward experience received by rewarded *S*s in the presence of others was extremely small when compared with the amount of daily contact our *S*s typically had with each other in the classroom, playground, and after school. (Practically all the *S*s in each of the classes had been together for their entire schooling, beginning with kindergarten.)

3. The choices made on the two pre-experimental sociometric tests (I and II) were found to be unexpectedly stable for third and fifth graders, indicating relatively reliable friendship ties[6] which the experimental experience was able to disrupt, however temporarily.[7]

Despite the above factors the prediction, that positive attitudes toward persons will be developed as a result of the receipt of reward in their presence, was clearly supported. Though the results were obtained with children, there is no reason to expect that they would not hold with adults as well. Such generalization must, of course, await adequate test.

The variable manipulated in the present study was simply reward vs. nonreward. Future research might profitably deal with variations in reward frequency, delay, and schedule, for example, as these affect the conditioning of attitudinal responses, and, consequently, as these affect the development of cohesiveness in small groups. That the results obtained in this study were predicted from general behavior principles increases our confidence in the promise which this kind of an approach holds for the general area of small group behavior.

Summary

Three-member groups of children played a game in which some members were rewarded and others were not. On a later sociometric test, outside of the game situation, rewarded *S*s chose a significantly greater proportion of their fellow group members than did the nonrewarded *S*s. These results were predicted from general principles of S-R learning theory.

REFERENCES

Bovard, E. W. The experimental production of interpersonal affect. *J. abnorm. soc. Psychol.*, 1951, **46**, 521–528.

Doob, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135–156.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Lott, Bernice E. Group cohesiveness: A learning phenomenon. *J. soc. Psychol.*, in press.

Thibaut, J. An experimental study of the cohesiveness of underprivileged groups. *Hum. Relat.*, 1950, **3**, 251–278.

(Received July 20, 1959)

---

[6] Forty percent of the *S*s made the same two choices on both tests; another 42% of the *S*s repeated one choice on both tests.

[7] No follow-up test was given since there was no reason to expect that the positive attitudes formed during the brief game situation would last, in the absence of continued reward experience in the presence of the same individuals.

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 301–310

THE OPEN SYSTEM IN PERSONALITY THEORY[1]

GORDON W. ALLPORT
Harvard University

OUR orb of common interest has two faces—one turned toward social psychology, the other toward personality. As things stand at the moment the first visage seems to me slightly depressed; the second slightly manic. However that may be, I should like to start this discussion by bringing the two faces into confrontation.

My first observation is that as members of Division 8 we are conspicuously the victims of fashion. Although our persona is sedate and seemly, we have our own hula hoops, flying saucers, and our own way of flagpole sitting. The interquartile range of our crazes I estimate to be about 10 years. McDougall's instinct theory held sway from 1908 to approximately 1920. Watsonian behaviorism dominated the scene for the next decade; then habit hierarchies, then field theory; now phenomenology. We never seem to solve our problems or exhaust our concepts; we only grow tired of them. At the moment it is fashionable to investigate response-set, coding, sensory deprivation, person perception, and to talk in terms of system theory—a topic to which we shall soon return. Ten years ago fashion called for group dynamics, Guttman scales, and for research on the centipedal properties of the authoritarian personality. Twenty years ago it was frustration-aggression, Thurstone scales, and national morale. And so it goes. We are fortunate that each surge of fashion leaves a rich residue of gain.

Fashions have their amusing and their serious side. We can smile at the way bearded problems receive tonsorial transformation. Having tired of "suggestibility" we adopt the new hairdo known as "persuasibility." Modern ethology excites us, and we are not troubled by the recollection that a century ago John Stuart Mill staked down the term to designate the new science of human character. We like the neurological concept of "gating," conveniently forgetting that American functionalism always stood firm for the dominance of general mental sets over specific. Reinforcement appeals to us but not the age-long debate over hedonism. The problem of freedom we brush aside in favor of "choice points." We avoid the body-mind problem but are in fashion when we talk about "brain models." Old wine, we find, tastes better from new bottles.

The serious side of the matter enters when we and our students forget that the wine is indeed old. Picking up a recent number of the *Journal of Abnormal and Social Psychology* I discover that the 21 articles written by American psychologists confine 90% of their references to publications of the past 10 years, although most of the problems they investigate have gray beards. In the same issue of the journal, three European authors locate 50% of their references prior to 1949. What this proves I do not know, excepting that European authors were not born yesterday. Is it any wonder that our graduate students reading our journals conclude that literature more than a decade old has no merit and can be safely disregarded? At a recent doctoral examination the candidate was asked what his thesis on physiological and psychological conditions of stress had to do with the body-mind problem. He confessed he had never heard of the problem. An undergraduate said that all he knew about Thomas Hobbes was that he sank with the *Leviathan* when it hit an iceberg in 1912.

A PSYCHOLINGUISTIC TRIFLE

Our windows are pretty much closed toward the past, but we rightly rejoice in our growth since World War II. Among the many fortunate developments is rejuvenation in the field of psycholinguistics. (Even here, however, I cannot refrain from pointing out that the much discussed Whorfian hypothesis was old stuff in the days of Wundt, Jespersen, and Sapir.) Be that as it may, I shall introduce my discussion of open systems in personality theory by a crude Whorfian analysis of our own vocabulary. My research (aided by the kind assistance of Stanley Plog) is too cursory to warrant detailed report.

---

[1] Invited address to Division 8—Personality and Social Psychology of the American Psychological Association, September 4, 1959.

What we did, in brief, was to study the frequency of the prefix *re-* and of the prefix *pro-* in psychological language. Our hypothesis was that *re-* compounds, connoting as they do again-ness, passivity, being pushed or maneuvered, would be far more common than *pro-* compounds connoting futurity, intention, forward thrust. Our sample consisted of the indexes of the *Psychological Abstracts* at 5-year intervals over the past 30 years; also all terms employing these prefixes in the Hinsie and Shatzky *Psychiatric Dictionary* and in the English and English *Psychological Dictionary*. In addition we made a random sampling of pages in five current psychological journals. Combining these sources it turns out that *re-* compounds are nearly five times as numerous as *pro-* compounds.

But, of course, not every compound is relevant to our purpose. Terms like reference, relationship, reticular, report do not have the connotation we seek; nor do terms like probability, process, and propaganda. Our point is more clearly seen when we note that the term reaction or reactive occurs hundreds of times, while the term proaction or proactive occurs only once—and that in English's *Dictionary*, in spite of the fact that Harry Murray has made an effort to introduce the word into psychological usage.

But even if we attempt a more strict coding of this lexical material, accepting only those terms that clearly imply reaction and response on one side and proaction or the progressive programing of behavior on the other, we find the ratio still is approximately 5:1. In other words our vocabulary is five times richer in terms like reaction, response, reinforcement, reflex, respondent, retroaction, recognition, regression, repression, reminiscence than in terms like production, proceeding, proficiency, problem solving, propriate, and programing. So much for the number of different words available. The disproportion is more striking when we note that the four terms reflex, reaction, response, and retention together are used 100 times more frequently than any single *pro-* compound excepting problem solving and projective—and this latter term, I submit, is ordinarily used only in the sense of reactivity.

The weakness of the study is evident. Not all terms connoting spontaneous, future oriented behavior begin with *pro-*. One thinks of expectancy, intention, purpose. But neither do all terms connoting passive responding or backward reference in time begin with *re-*. One thinks of coding, traces, input-output, and the like. But while our analysis leaves much to be desired it prepares the way for our critique of personality theory in terms of systems. The connecting link is the question whether we have the verbal, and therefore the conceptual, tools to build a science of change, growth, futurity, and potential; or whether our available technical lexicon tends to tie us to a science of response, reaction, and regression. Our vocabulary points to personality development from the past up to now more readily than to its development from here-on-out into the future.

THE CONCEPT OF SYSTEM

Until a generation or so ago science, including psychology, was preoccupied with what might be called "disorganized complexity." Natural scientists explored this fragment and that fragment of nature; psychologists explored this fragment and that fragment of experience and behavior. The problem of interrelatedness, though recognized, was not made a topic for direct inquiry.

What is called system theory today—at least in psychology—is the outgrowth of the relatively new organismic conception reflected in the work of von Bertalanffy, Goldstein, and in certain aspects of gestalt psychology. It opposes simple reaction theories where a virtual automaton is seen to respond discretely to stimuli as though they were pennies-in-the-slot. Interest in system theory is increasing in psychology, though perhaps not so fast as in other sciences.

Now a system—any system—is defined merely as a complex of elements in mutual interaction. Bridgman (1959), as might be expected of an operationist, includes a hint of method in his definition. He writes, a system is "an isolated enclosure in which all measurements that can be made of what goes on in the system are in some way correlated" (p. 188).

Systems may be classified as closed or open. A closed system is defined as one that admits no matter from outside itself and is therefore subject to entropy according to the second law of thermodynamics. While some outside energies, such as change in temperature and wind may play upon a closed system, it has no restorative properties and no transactions with its environment, so that like a decaying bridge it sinks into thermodynamic equilibrium.

Now some authors, such as von Bertalanffy (1952b), Brunswik (1955), and Pumpian-Mindlin (1959), have said or implied that certain theories of psychology and of personality operate with the conception of closed systems. But in my opinion these critics press their point too far. We had better leave closed systems to the realm of physics where they belong (although even here it is a question whether Einstein's formula for the release of matter into energy does not finally demonstrate the futility of positing a closed cgs system even in physics). In any event it is best to admit that all living organisms partake of the character of open systems. I doubt that we shall find any advocate of a truly closed system in the whole range of personality theory. At the same time current theories do differ widely in the amount of openness they ascribe to the personality system.

If we comb definitions of open systems we can piece together four criteria: there is intake and output of both matter and energy; there is the achievement and maintenance of steady (homeostatic) states, so that the intrusion of outer energy will not seriously disrupt internal form and order; there is generally an increase of order over time, owing to an increase in complexity and differentiation of parts; finally, at least at the human level, there is more than mere intake and output of matter and energy: there is extensive transactional commerce with the environment.[2]

While all of our theories view personality as an open system in some sense, still they can be fairly well classified according to the varying emphasis they place upon each of these criteria, and according to how many of the criteria they admit.

Criterion 1

Consider the first criterion of material and energy exchange. Stimulus-response theory in its purest form concentrates on this criterion to the virtual exclusion of all the others. It says in effect that a stimulus enters and a response is emitted. There is, of course, machinery for summation, storage, and delay, but the output is broadly commensurate with the intake. We need study only the two poles of stimulus and response with a minimum of concern for intervening processes. Methodological positivism goes one step further, saying in effect, that we do not need the concept of personality at all. We focus attention on our own measurable manipulations of input and on the measurable manipulations of output. Personality thus evaporates in a mist of method.

Criterion 2

The requirement of steady state for open systems is so widely admitted in personality theory that it needs little discussion. To satisfy needs, to reduce tension, to maintain equilibrium, comprise, in most theories, the basic formula of personality dynamics. Some authors, such as Stagner (1951) and Mowrer (1959) regard this formula as logically fitting in with Cannon's (1932) account of homeostasis.[3] Man's intricate adjustive behavior is simply an extension of the principle involved in temperature regulation, balance of blood volume, sugar content, and the like, in the face of environmental change. It is true that Toch and Hastorf (1955) warn against overextending the concept of homeostasis in personality theory. I myself doubt that Cannon would approve the extension, for to him the value of homeostasis lay in its capacity to free man for what he called "the priceless unessentials" of life (1932, p. 323). When biological equilibrium is attained the priceless unessentials take over and constitute the major part of human activity. Be that as it may, most current theories clearly regard personality as a *modus operandi* for restoring a steady state.

---

[2] von Bertalanffy's definition explicitly recognizes the first two of these criteria as present in all living organisms. A living organism, he says, is "an open system which continually gives up matter to the outer world and takes in matter from it, but which maintains itself in this continuous exchange in a steady state, or approaches such steady state in its variations in time" (1952a, p. 125). But elsewhere in this author's writing we find recognition of the additional criteria (1952a, p. 145; 1952b, p. 34).

[3] In a recent review Mowrer (1959) strongly defends the homeostatic theory. He is distressed that the dean of American psychologists, Robert Woodworth (1958) has taken a firm stand against the "need-primacy" theory in favor of what he calls the "behavior-primacy" theory. With the detailed merits of the argument we are not here concerned. What concerns us at the moment is that the issue has been sharply joined. Need-primacy which Mowrer calls a "homeostatic" theory does not go beyond our first two criteria for an open system. Woodworth by insisting that contact with, and mastery of, the environment constitute a pervasive principle of motivation, recognizes the additional criteria.

Psychoanalytic theories are of this order. According to Freud the ego strives to establish balance among the three "tyrants"—id, superego, and outer environment. Likewise the so-called mechanisms of ego defense are essentially maintainers of a steady state. Even a neurosis has the same basic adjustive function.[4]

To sum up: most current theories of personality take full account of two of the requirements of an open system. They allow interchange of matter and energy, and recognize the tendency of organisms to maintain an orderly arrangement of elements in a steady state. Thus they emphasize stability rather than growth, permanence rather than change, "uncertainty reduction" (information theory), and "coding" (cognitive theory) rather than creativity. In short, they emphasize being rather than becoming. Hence, most personality theories are biologistic in the sense that they ascribe to personality only the two features of an open system that are clearly present in all living organisms.

There are, however, two additional criteria, sometimes mentioned but seldom stressed by biologists themselves, and similarly neglected in much current personality theory.

Transatlantic Perspective

Before examining Criterion 3 which calls attention to the tendency of open systems to enhance their degree of order, let us glimpse our present theoretical situation in cross-cultural perspective. In this country our special field of study has come to be called "behavioral science" (a label now firmly stuck to us with the glue of the Ford millions). The very flavor of this term suggests that we are occupied with semiclosed systems. By his very name the "behavioral scientist" seems committed to study man more in terms of behavior than in terms of experience, more in terms of mathematical space and clock-time than in terms of existential space and time; in terms of response more than in terms of programing; in terms of tension reduction more than tension enhancement; in terms of reaction more than proaction.

Now let us leap our cultural stockade for a moment and listen to a bit of ancient Hindu wisdom. Most men, the Hindus say, have four central desires. To some extent, though only roughly, they correspond to the developmental stages of life. The first desire is for pleasure—a condition fully and extensively recognized in our Western theories of tension reduction, reinforcement, libido, and needs. The second desire is for success—likewise fully recognized and studied in our investigations of power, status, leadership, masculinity, and need-achievement. Now the third desire is to do one's duty and discharge one's responsibility. (It was Bismarck, not a Hindu, who said: "We are not in this world for pleasure but to do our damned duty.") Here our Western work begins to fade out. Excepting for some pale investigations of parental punishment in relation to the development of childhood conscience, we have little to offer on the "duty motive." Conscience we tend to regard as a reactive response to internalized punishment, thus confusing the past "must" of learning with the "ought" involved in programing our future (Allport, 1954, pp. 68–74). Finally, the Hindus tell us that in many people all these three motives pall, and they then seek intensely for a grade of understanding—for a philosophical or religious meaning—that will liberate them from pleasure, success, and duty (Smith, 1958). (Need I point out that most

---

[4] When we speak of the "function" of a neurosis we are reminded of the many theories of "functionalism" current in psychology and social science. Granted that the label, as Merton (1957) has shown, is a wide one, still we may safely say that the emphasis of functionalism is always on the usefulness of an activity in maintaining the "steady state" of a personality or social or cultural system. In short, "functional" theories stress maintenance of present direction allowing little room or none at all for departure and change.

Western personality theories treat the religious aspiration in reactive terms as an escape device, to be classified along with suicide, alcoholism, and neurosis?)

Now we retrace our steps from India to modern Vienna and encounter the existentialist school of logotherapy. Its founder, Viktor Frankl, emphasizes above all the central place of duty and meaning, the same two motives that the Hindus place highest in their hierarchy of desire. Frankl reached his position after a long and agonizing incarceration in Nazi concentration camps. With other prisoners he found himself stripped to naked existence (1959a). In such extremity what does a person need and want? Pleasure and success are out of the question. One wants to know the meaning of his suffering and to learn how as a responsible being he should acquit himself. Should he commit suicide? If so, why; if not, why not? The search for meaning becomes supreme.

Frankl is aware that his painfully achieved theory of motivation departs widely from most American theory, and he points out the implication of this fact for psychotherapy. He specifically criticizes the principle of homeostasis (1959b) as implying that personality is a quasiclosed system. To cater to the internal adjustments of a neurotic, or to assume that he will regain health by reshuffling his memories, defenses, or conditioned reflexes is ordinarily self-defeating. In many cases of neurosis only a total breakthrough to new horizons will turn the trick.

Neither Hindu psychology nor Frankl underestimates the role of pleasure and success in personality. Nor would Frankl abandon the hard won gains reflected in psychoanalytic and need theory. He says merely that in studying or treating a person we often find these essentially homeostatic formulations inadequate. A man normally wants to know the whys and wherefores. No other biological system does so; hence, man stands alone in that he possesses a degree of openness surpassing that of any other living system.

Criterion 3

Returning now to our main argument, we encounter a not inconsiderable array of theories that emphasize the tendency of human personality to go beyond steady states and to strive for an enhancement and elaboration of internal order even at the cost of considerable disequilibrium.

I cannot examine all of these nor name all the relevant authors. One could start with McDougall's proactive sentiment of self-regard which he viewed as organizing all behavior through a kind of "forward memory" (to use Gooddy's apt term—1959). Not too dissimilar is the stress that Combs and Snygg place on the enhancement of the phenomenal field. We may add Goldstein's conception of self-actualization as tending to enhance order in personality; also Maslow's theory of growth motives that supplement deficiency motives. One thinks of Jung's principle of individuation leading toward the achievement of a self (a goal never actually completed). Some theories, Bartlett and Cantril among them, put primary stress on the "pursuit of meaning." Certain developments in post-Freudian "ego psychology" belong here.[5] So too does existentialism with its recognition of the need for meaning and of the values of commitment. (The brain surgeon, Harvey Cushing, was speaking of open systems when he said: "The only way to endure life is to have a task to complete.")

No doubt we should add Woodworth's recent advocacy of the "behavior primacy" theory as opposed to the "need" theory, Robert White's emphasis on "competence," and Erikson's "search for identity."

Now these theories are by no means identical. The differences between them merit prolonged debate. I lump them here simply because all seem to me to recognize the third criterion of open systems, namely, the tendency of such systems to enhance their degree of order and become something more than at present they are.

We all know the objection to theories of this type. Methodologists with a taste for miniature and fractionated systems complain that they do not lead to "testable propositions" (cf. Roby, 1959). The challenge is valuable in so far as it calls for an expansion of research ingenuity. But the complaint is ill-advised if it demands that we return to quasiclosed systems simply because they are more "researchable" and elegant. Our task is to study what is, and not what is immediately convenient.

---

[5] Pumpian-Mindlin (1959) writes: "The focus of clinical psychoanalysis on ego psychology is a direct result of the change from a closed system to an open one" (p. 1051).

Criterion 4

Now for our fourth and last criterion. Virtually all the theories I have mentioned up to now conceive of personality as something integumented, as residing within the skin. There are theorists (Kurt Lewin, Martin Buber, Gardner Murphy, and others) who challenge this view, considering it too closed. Murphy says that we overstress the separation of man from the context of his living. Experiments on sensory deprivation Hebb (1955) has interpreted as demonstrations of the constant dependence of inner stability on the flow of environmental stimulation. Why Western thought makes such a razor-sharp distinction between the person and all else is an interesting problem. Probably the personalistic emphasis in Judeo-Christian religion is an initial factor, and as Murphy (1958, p. 297) has pointed out the industrial and commercial revolutions further accentuated the role of individuality. Shinto philosophy, by contrast, regards the individual, society, and nature as forming the tripod of human existence. The individual as such does not stick out like a raw digit. He blends with nature and he blends with society. It is only the merger that can be profitably studied.

As Western theorists most of us, I dare say, hold the integumented view of the personality system. I myself do so. Others rebelling against the setting of self over against the world, have produced theories of personality written in terms of social interaction, role relations, situationism, or some variety of field theory. Still other writers, such as Talcott Parsons (1951) and F. H. Allport (1955), have admitted the validity of both the integumented personality system and systems of social interaction, and have spent much effort in harmonizing the two types of systems thus conceived.

This problem, without doubt, is the knottiest issue in contemporary social science. It is the issue which, up to now, has prevented us from agreeing on the proper way to reconcile psychological and sociocultural science.

In this matter my own position is on the conservative side. It is the duty of psychology, I think, to study the person-system, meaning thereby the attitudes, abilities, traits, trends, motives, and pathology of the individual—his cognitive styles, his sentiments, and individual moral nature and their interrelations. The justification is twofold: (*a*) there is a persistent though changing person-system in time, clearly delimited by birth and death; (*b*) we are immediately aware of the functioning of this system; our knowledge of it, though imperfect, is direct, whereas our knowledge of all other outside systems, including social systems, is deflected and often distorted by their necessary incorporation into our own apperceptions.

At the same time our work is incomplete unless we admit that each person possesses a *range* of abilities, attitudes, and motives that will be evoked by the different environments and situations he encounters. Hence, we need to understand cultural, class, and family constellations and traditions in order to know the schemata the person has probably interiorized in the course of his learning. But I hasten to warn that the study of cultural, class, family, or any other social system does not automatically illumine the person-system, for we have to know whether the individual has accepted, rejected, or remained uninfluenced by the social system in question. The fact that one plays the role of, say, teacher, salesman, or father is less important for the study of his personality than to know whether he likes or dislikes, and how he defines, the role. And yet at the same time unless we are students of sociocultural systems we shall never know what it is the person is accepting, rejecting, or redefining.

The provisional solution I would offer is the following: the personality theorist should be so well trained in social science that he can view the behavior of an individual as fitting any system of interaction; that is, he should be able to cast this behavior properly in the culture where it occurs, in its situational context, and in terms of role theory and field theory. At the same time he should not lose sight—as some theorists do—of the fact that there is an internal and subjective patterning of all these contextual acts. A traveler who moves from culture to culture, from situation to situation, is none the less a single person;

and within him one will find the nexus, the patterning, of the diverse experiences and memberships that constitute his personality.

Thus, I myself would not go so far as to advocate that personality be defined in terms of interaction, culture, or roles. Attempts to do so seem to me to smudge the concept of personality, and to represent a surrender of the psychologist's special assignment as a scientist. Let him be acquainted with all systems of interaction, but let him return always to the point where such systems converge and intersect and are patterned—in the single individual.

Hence, we accept the fourth (transactional) criterion of the open system, but with the firm warning that it must not be applied with so much enthusiasm that we lose the personality system altogether.

GENERAL SYSTEMS THEORY

There are those who see hope for the unification of science in what James Miller (1955) called "general behavior systems theory." This approach seeks formal identities between physical systems, the cell, the organ, the personality, small groups, the species, and society. Now critics (e.g., Buck, 1956) complain that all this is feeble analogizing, that formal identities probably do not exist, and that attempts to express analogies in terms of mathematical models result only in the vaguest generalities. As I see it, the danger in attempting to unify science in this manner lies in the inevitable approach from below, that is, in terms of physical and biological science. Closed systems or systems only partly open become our model, and if we are not careful, human personality in all its fullness is taken captive into some autistic paradise of methodology.

Besides neglecting the criteria of enhanced organization and transaction general systems theory has an added defect. The human person is, after all, the observer and interpreter of systems. This awkward fact has recently been haunting the founder of the operational movement, P. W. Bridgman (1959). Can we as scientists live subjectively within our system and at the same time take a valid objective view thereof?

Some years ago Elkin (1947) published the case of "Harry Holzer," and invited 39 specialists to offer their conceptualizations. As might be expected, many different conceptualizations resulted. No theorist was able entirely to divest the case of his own preconceptions. Each read the objective system in terms of the subjective. In other words, our theories of personality—all of them—reflect the temperament of the author fully as much as the personality of *alter*.

This sad spectre of observer contamination should not, I think, discourage us from the search for objectively valid theory. Truth, as the philosopher Charles Pierce has said, is the opinion which is fated to be ultimately agreed to by all who investigate. My point is that "the opinion fated to be ultimately agreed to by all who investigate" is not likely to be reached through a premature application of general systems theory, nor through devotion to any one partially closed theory. Theories of open-systems hold more promise, though at present they are not in agreement among themselves. But somewhere, sometime, I hope and believe, we shall establish a theory of the nature of personality which all wise men who investigate, including psychologists, will eventually accept.

SOME EXAMPLES

In the meantime, I suggest that we regard all sharp controversies in personality theory as probably arising from the two opposed points of view—the quasiclosed and the fully open.

The principle of reinforcement, to take one example, is commonly regarded as the cement that stamps in a response, as the glue that fixes personality at the level of past deeds. Now an open-system interpretation is very different. Feigl (1959, p. 117), for instance, has pointed out that reinforcement works primarily in a prospective sense. It is only from a *recognition* of consequences (not from the consequences themselves) that the human individual binds the past to the future and resolves to avoid punishment and to seek rewards in similar circumstances, provided, of course, that it is consonant with his interests and values to do so. Here we no longer assume that reinforcement stamps in, but that it is one factor among many to be considered in the programing of future action (Allport, 1946). In this example we see what a wide difference it makes whether we regard personality as a quasiclosed or open system.

The issue has its parallels in neurophysiology. How open is the nervous system? We know it is of a complexity so formidable that we have only an inkling as to how complex it may be. Yet one thing is certain, namely, that high level gating often controls and steers lower level processes. While we cannot tell exactly what we mean by "higher levels" they surely involve ideational schemata, intentions, and generic personality trends. They are instruments for programing, not merely for reacting. In the future we may confidently expect that the neurophysiology of programing and the psychology of proaction will draw together. Until they do so it is wise to hold lightly our self-closing metaphors of sowbug, switchboard, giant computor, and hydraulic pump.

Finally, an example from motivation theory. Some years ago I argued that motives may become functionally autonomous of their origins. (And one lives to regret one's brashness.)

Whatever its shortcomings the concept of functional autonomy succeeds in viewing personality as an open and changing system. As might be expected, criticism has come chiefly from those who prefer to view the personality system as quasiclosed. Some critics say that I am dealing only with occasional cases where the extinction of a habit system has failed to occur. This criticism, of course, begs the question, for the point at issue is why do some habit systems fail to extinguish when no longer reinforced? And why do some habit systems that were once instrumental get refashioned into interests and values having a motivational push?

The common counterargument holds that "secondary reinforcement" somehow miraculously sustains all the proactive goal-seeking of a mature person. The scientific ardor of Pasteur, the religio-political zeal of Gandhi, or for that matter, Aunt Sally's devotion to her needlework, are explained by hypothetical cross-conditioning that somehow substitutes for the primary reinforcement of primary drives. What is significant for our purposes is that these critics prefer the concept of secondary reinforcement, not because it is clearer, but because it holds our thinking within the frame of a quasiclosed (reactive) system.

Now is not the time to re-argue the matter, but I have been asked to hint at my present views. I would say first that the concept of functional autonomy has relevance even at the level of quasiclosed systems. There are now so many indications concerning feedback mechanisms, cortical self-stimulation, self-organizing systems, and the like (Chang, 1950; Hebb, 1949; Olds & Milner, 1954) that I believe we cannot deny the existence of self-sustaining circuit mechanisms which we can lump together under the rubric "perseverative functional autonomy."

But the major significance of the concept lies in a different direction, and presupposes the view that personality is an expanding system seeking progressively new levels of order and transaction. While drive motives remain fairly constant throughout life, existential motives do not. It is the very nature of an open system to achieve progressive levels of order through change in cognitive and motivational structure. Since in this case the causation is systemic we cannot hope to account for functional autonomy in terms of specific reinforcements. This condition I would call "propriate functional autonomy."

Both perseverative and propriate autonomy are, I think, indispensable conceptions. The one applies to the relatively closed part-systems within personality; the other to the whole continuously evolving structure of the whole.

A last example. It is characteristic of the quasiclosed system outlook that it is heavily nomothetic. It seeks response and homeostatic similarities among all personality systems (or, as in general behavior systems theory, among *all* systems). If, however, we elect the open system view we find ourselves forced in part toward the idiographic outlook. For now the vital question becomes "what makes the system hang together in any one person?" (cf. Taylor, 1958). Let me repeat this question, for it is the one that more than any other has haunted me over the years. *What makes the system cohere in any one person?* That this problem is pivotal, urgent, and relatively neglected, will be recognized by open-system theorists, even while it is downgraded and evaded by those who prefer their systems semiclosed.

FINAL WORD

If my discourse has seemed polemical I can only plead that personality theory lives by controversy. In this country we are fortunate that no single party line shackles our speculations. We are free to pursue any and all assumptions concerning the nature of man. The penalty we pay is that for the present we cannot expect personality *theory* to be cumulative—although, fortunately, to some extent personality *research* can be.

Theories, we know, are ideally derived from axioms, and if axioms are lacking, as in our field they are, from assumptions. But our assumptions regarding the nature of man range from the Adlerian to the Zilborgian, from the Lockean to the Leibnitzian, from the Freudian to the Hullian, and from the cybernetic to the existentialist. Some of us model man after the pigeon; others view his potentialities as many splendored. And there is no agreement in sight.

Nils Bohr's principle of complementarity contains a lesson for us. You recall that he showed that if we study the position of a particle we cannot at the same time study its momentum. Applied to our own work the principle tells us that if we focus on reaction we cannot simultaneously study proaction; if we measure one trait we cannot fix our attention on pattern; if we tackle a subsystem we lose the whole; if we pursue the whole we overlook the part-functioning. For the single investigator there seems to be no escape from this limitation. Our only hope is to overcome it by a complementarity of investigators and of theorists.

While I myself am partisan for the open system, I would shut no doors. (Some of my best friends are quasiclosed systematists.) If I argue for the open system I plead more strongly for the open mind. Our condemnation is reserved for that peculiar slavery to fashion which says that conventionality alone makes for scientific respectability. We still have much to learn from our creative fumblings with the open system. Among our students, I trust, there will be many adventurers. Shall we not teach them that in the pastures of science it is not only the sacred cows that can yield good scientific milk?

REFERENCES

ALLPORT, F. H. *Theories of personality and the concept of structure.* New York: Wiley, 1955.

ALLPORT, G. W. Effect: A secondary principle of learning. *Psychol. Rev.*, 1946, **53**, 335–347.

ALLPORT, G. W. *Becoming: Basic considerations for a psychology of personality.* New Haven: Yale Univer. Press, 1954.

BRIDGMAN, P. W. *The way things are.* Cambridge: Harvard Univer. Press, 1959.

BRUNSWIK, E. The conceptual framework of psychology. In, *International Encyclopedia of Unified Science.* Vol. 1. Chicago: Chicago Univer. Press, 1955.

BUCK, R. C. On the logic of general behavior systems theory. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science.* Vol. 1. Minneapolis: Univer. Minnesota Press, 1956.

CANNON, W. B. *The wisdom of the body.* New York: Norton, 1932.

CHANG, H. T. The repetitive discharge of corticothalamic reverberating circuit. *J. Neurophysiol.*, 1950, **13**, 235–257.

ELKIN, F. Specialists interpret the case of Harry Holzer. *J. abnorm. soc. Psychol.*, 1947, **42**, 99–111.

FEIGL, H. Philosophical embarrassments of psychology. *Amer. Psychologist*, 1959, **14**, 115–128.

FRANKL, V. E. *From death camp to existentialism.* Boston: Beacon, 1959. (a)

FRANKL, V. E. Das homöostatische Prinzip und die dynamische Psychologie. *Z. Psychother. med. Psychol.*, 1959, **9**, 41–47. (b)

GOODDY, W. Two directions of memory. *J. indiv. Psychol.*, 1959, **15**, 83–88.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HEBB, D. O. The mammal and his environment. *Amer. J. Psychiat.*, 1955, **111**, 826–831. (Reprinted: In Eleanor Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1958. Pp. 335–341.)

MERTON, R. K. *Social theory and social structure.* (Rev. ed.) Glencoe, Ill.: Free Press, 1957.

MILLER, J. G. Toward a general theory for the behavioral sciences. *Amer. Psychologist*, 1955, **10**, 513–531.

MOWRER, O. H. A cognitive theory of dynamics. Review of R. S. Woodworth, *Dynamics of behavior. Contemp. Psychol.*, 1959, **4**, 129–133.

MURPHY, G. *Human potentialities.* New York: Basic Books, 1958.

OLDS, J., & MILNER, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. comp. physiol. Psychol.*, 1954, **47**, 419–427.

PARSONS, T. *The social system.* Glencoe, Ill.: Free Press, 1951.

PUMPIAN-MINDLIN, E. Propositions concerning energetic-economic aspects of libido theory. *Ann. NY Acad. Sci.*, 1959, **76**, 1038–1052.

ROBY, T. B. An opinion on the construction of behavior theory. *Amer. Psychologist*, 1959, **14**, 127–134.

SMITH, H. *The religions of man.* New York: Harper, 1958. (Mentor Book Ed., 1959.)

STAGNER, R. Homeostasis as a unifying concept in personality theory. *Psychol. Rev.*, 1951, **58**, 5–17.

TAYLOR, J. G. Experimental design: A cloak for intellectual sterility. *Brit. J. Psychol.*, 1958, **49**, 106–116.

TOCH, H. H., & HASTORF, A. H. Homeostasis in psychology. *Psychiatry*, 1955, **18**, 81–91.

VON BERTALANFFY, L. *Problems of life.* (Trans. of *Das biologische Weltbild*, 1949) New York: Wiley, 1952. (a)

VON BERTALANFFY, L. Theoretical models in biology and psychology. In D. Krech & G. S. Klein (Eds.), *Theoretical models and personality theory.* Durham: Duke Univer. Press, 1952. (b)

WOODWORTH, R. S. *Dynamics of behavior.* New York: Holt, 1958.

(Received September 9, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 311-318

ON THE RELIABILITY OF THE ANAMNESTIC INTERVIEW[1]

ERNEST A. HAGGARD[2]
College of Medicine, University of Illinois

ARNE BREKSTAD AND ÅSE GRUDA SKARD
Psychological Institute, University of Oslo

THE anamnestic interview is used extensively by psychiatrists, guidance counselors, anthropologists, and other persons who are interested in the prior events in the life of another person. This procedure, which relies upon the memory of an informant, is used so often simply because the easiest way (and often the only way) to obtain many types of information about another person's background is to ask him. And, because the preoccupation of the interviewer is to obtain "the truth," he generally is inclined to accept with gratitude whatever the informant tells him, and to reify it.

Many investigators have questioned the extent to which anamnestic material gives a true picture of prior events. Although the vast literature on memory and its fallibility bears on this question, relatively few studies have been carried out that deal with it directly. The purpose of this study is to inquire about the reliability of various types of anamnestic information and to throw light on some of the factors that may distort an informant's report during an anamnestic interview.

METHOD

Data and sample. The data are based on the records of a longitudinal study of the families of 19 children born and reared in Oslo and on a follow-up anamnestic interview with the mothers of these children. Over the duration of this study numerous interview and observational data have been collected on the parents and, in addition, many types of measurement, test, and observational data have been collected on the children. These data were collected for several purposes, one of which was to provide the opportunity for later investigating the reliability of the anamnestic interview.[3]

The sample was drawn from 55 working class families who visited a well-baby clinic in Oslo in 1952–53. Of the 19 families, 12 were to have their first child (4 were boys, 8 were girls) and 7 were to have their second child (6 were boys, 1 was a girl). Since the parents were interviewed before the children were born and at successive intervals thereafter, it was possible to sample several time periods and several areas of information or types of scales. The children were between 7 and 8 years of age at the time of the anamnestic interview.

Sampling time periods. Three time samples were selected for study: the prenatal period (about a month before delivery), during the child's first year, and during the child's sixth year. Thus, whereas the prenatal scales have to do primarily with the parents' projected attitudes and expectations, the first and sixth year scales are tied more closely to their experiences with their child and what they expected from him at that age.

From 18 scales available from the larger project for the prenatal period, 10 were selected; from 31 scales available for the first year period, 6 were selected; and from 36 scales available for the sixth year period, 11 were selected. The 27 scales used in this study are given in Table 1 below.

With respect to the possible effect of time on the reliability of memory, two samples of scales covered an interval of from 1 to 2 years and two samples of scales covered an interval of from 7 to 8 years. For the former scales, one sample (Scales 1–3) covered an interval of about 1 year (these three scales were readministered during the first year and were anamnestic to the prenatal period); the other sample (Scales 17–27) covered an interval of from 1 to 2 years. For the longer intervals, one sample (Scales 1–10) covered an interval of about 8 years and the other sample (Scales 11–16) covered an interval of about 7 years (see Table 1).

Sampling types of information (scales). Five criteria were used to select the scales for this study: (*a*) that, where possible, they should be representative of the questions which guidance counselors, clinical psychologists, etc. typically ask mothers in an anamnestic interview; (*b*) that it would not be unreasonable to ask the mothers for such information in view of their previous contacts with the project; (*c*) that the scales should cover a variety of types of information; (*d*) that the project records should contain sufficiently consistent and precise information on the 19 cases to minimize uncertainty in our scaling procedure;[4] and (*e*) where possible, the scales should not be highly intercorrelated.[5]

---

[1] This study was supported by Research Grant M-1737 from the National Institute of Mental Health, of the National Institutes of Health, United States Public Health Service.

[2] On leave from the Department of Psychiatry, University of Illinois during 1959–60 as a Public Health Service Research Fellow of the National Institute of Mental Health.

[3] This project, entitled "Children's Personality Development in an Oslo Milieu," was begun in 1950 under the direction of Åse G. Skard. A full description of this project is available in: Skard, Åse G. Scope and methods of the Oslo project. 1960. (Mimeo)

Scaling procedure. After examining the protocols, we decided to use a normalized six-point scale to quantify the data. For convenience, we will call the scales based on the project protocols "historical" and the corresponding scales based on the follow-up interview "anamnestic."

Although the historical scales for the prenatal and first year periods were originally in the form of ratings, it was necessary to recode them to have scales of uniform size and explicit definition. The historical scales for the sixth year period were originally in the form of ranks based on the independent judgments previously made by two other raters.[6] Another difference between the earlier and later scales lies in the fact that the prenatal and first year scales were generally based on specific questions to the mothers, whereas the sixth year scales were more often inferred from the "general tone" of the historical material. For the sixth year period, then, it was not possible to have completely parallel scoring procedures, since in the anamnestic interview, the questions were phrased to bear directly on Scales 17 to 27.

For the historical scales, the extremes and midpoints were anchored by excerpts from the protocols, the cases coded along a semiquantitative graphic continuum (see Champney, 1941), and later converted into a normalized six-point scale. The same person who established the mid- and end-points on the historical scales also coded the anamnestic scales. Each historical scale was coded also in terms of whether the mother had shown marked anxiety in connection with it at the time. From four to eight mothers were sufficiently anxious in connection with the content of Scales 1, 4, 5, and 7 to permit later analysis of the possible effects of such disturbances on the reliability of recall.

In coding the anamnestic scales we wished to obtain some measure of the extent of any random discrepancies or systematic shifts in the mothers' responses to the questions corresponding to each of the scales. Thus, on each anamnestic scale, each mother's response was coded in terms of where it would have fallen on the corresponding historical scale. Because discrepancies did occur between the mothers' reports in the historical and anamnestic interviews, 4 of the 513 scores on the anamnestic scales fell beyond the 1–6 point range. In these instances we decided to use the scale value of 1 or 6 to preserve the uniform size of the scales. Also, because of changes or shifts in the mothers' reports, not all of the anamnestic scales were distributed normally, although the distribution over all of the scales was approximately normal.

The anamnestic interview. The interview itself was justified to the mothers in terms of our interest in "how they see things now" and our desire to learn how the children were doing in school (since, in Norway, children begin school in their seventh year). Although the mothers had been interviewed many times by members of the project staff, they were not acquainted with the present interviewer.

An interview guide was drawn up around the 27 scales with the addition of questions to fill in time gaps and to cover various points of interest to the project staff. Although this interview, which usually lasted a little over 1 hour, was designed to be a representative anamnestic interview, it was atypical in three respects: it was recorded; the mothers knew that the project files contained their previous responses to similar questions; and the mothers did not come to the interview because they were disturbed about, and hence seeking help for, some "problem." If these three differences have introduced any bias into our findings, their effect has probably been to elicit more accurate or reliable responses than would have been the case if the mothers had been emotionally disturbed or had had reason to give an unduly favorable (or unfavorable) account to the interviewer.

Methods of data analysis. The coefficient of intraclass correlation, *R*, was used as the most sensitive and accurate measure of reliability (see Haggard, 1958) to reflect the relative congruence of the mothers' responses to the questions underlying the historical and anamnestic scales. Appropriate coefficients were computed over scales and over individuals, and frequently these *R*s were averaged (via the *z* transformation) to provide measures for groups of scales or individuals.

In analyzing these data, our primary interest has been to examine the extent to which various factors or conditions have contributed to the reliable (or distorted) recall in an anamnestic interview and to indicate their relative importance rather than to stress the statistical significance of our findings or to infer that such factors will have comparable influence in all similar contexts.

RESULTS

Validity or Reliability?

In order to speak of validity in the customary sense, one would need objective external criteria for the scales. Such criteria were available for Scales 11–14 since each child's length, weight, and time of birth were available from hospital records and the approximate length of breast-feeding was available from clinic records and frequent visits to the home during the first year. The *R*s for these four scales ranged

---

[4] For one case, which had dropped out of the study, 3's and 4's were assigned at random for the 11 sixth year scales (17–27). Otherwise, the protocols contained adequate information for assigning the remaining 502 (of the 513) historical scores.

[5] For the sixth year period, for example, the rank order intercorrelations among the 11 scales ranged from $-.39$ to $.55$, with a median ρ of $.03$.

[6] The interjudge ρ's ranged from .50 to .88 before the two rankings were combined to form the basis for Scales 17–27.

TABLE 1
RELIABILITY COEFFICIENTS BETWEEN HISTORICAL AND ANAMNESTIC REPORTS

Prenatal scales:	Six-month anamnesis	Eight-year anamnesis
1. Attitude toward being pregnant	.02	.60
2. Strength of wish for a certain sex	.50	.23
3. What sex did the mother wish for	.81	.59
4. Did the parents plan to have the child		.83
5. Pregnancy difficulties		.40
6. Anxiety for the childbirth		.09
7. Anxiety for the child's normality and appearance		.32
8. Attitude toward regulating child's food schedule		.41
9. Attitude toward obedience		.37
10. Attitude toward punishment as discipline		.73
		$R_{av.}$ = .49
First year scales:	R from measures	
11. Length of child at birth	.91	.89
12. Weight of child at birth	.77	.71
13. Child born at normal time	.91	.81
14. Length of breast-feeding	.96	.77
15. Attitude toward the delivery (childbirth)		.56
16. Attitude toward regulating child's food schedule		.46
		$R_{av.}$ = .73
Sixth year scales:		
17. Security (or lack of anxiety) in child care		.49
18. Agreement between the parents in child care		.22
19. Encouraging the child to dress alone		.13
20. Encouraging the child to wash himself alone		.35
21. Encouraging the child to eat alone		.25
22. Encouraging the child to go to the toilet alone		.17
23. Giving child freedom to leave home alone		.53
24. Letting child decide things by himself		−.09
25. Attitude toward aggression in general		.35
26. Attitude toward child's aggression to other children		.18
27. Letting child choose his own playmates		−.08
		$R_{av.}$ = .24
		$R_{av.}$ (total interview) = .47

from .71 to .89, with an average R ($R_{av.}$) of .80.[7] It is worth noting that this group of four scales contains the highest Rs in Table 1.

Even though the interviewer may be seeking "the truth" in conducting an anamnestic interview, in most cases it is meaningless from a research point of view to phrase the matter in these terms, simply because there is no way of telling what was "true" at the time the event in question took place. One could, of course, assume that the informant's report at the time should be taken at face value but, even so, this assumption would hardly enable one to speak of "validity." With respect to Scale 6, for example, the mother stated on a given occasion how anxious she was about the coming childbirth—but who knows whether she was also trying to be brave or to seek sympathy, or whether her statements on another day (and in a different frame of mind) would have been the same?

Although we may speak of the findings for Scales 11–14 as validity coefficients, on the remaining 23 scales we will limit our interest to the degree of congruence between the historical and anamnestic reports, as indicated by Rs, in order to examine various types of factors associated with the degree of agreement between these two sets of reports. More specifically, we will focus our attention on four types of conditions that appear to influence the reliability of the mothers' reports during the anamnestic interview. These conditions involve the length of the time interval between the historical and anamnestic interviews, differences in the type of information underlying the scales, sex and birth-order differences among the children, and individual differences among the mothers.

Time Intervals

Psychologists ordinarily assume that more distant events will be recalled less well than more recent events—with some exceptions, of course. In this study it was possible to group the 27 scales into those which involved a relatively short time interval (1 to 2 years) and those which involved a relatively long interval (7 to 8 years). There were two samples of scales in each grouping as follows:

1. Short interval: prenatal to first year. Scales 1–3 were based on questions asked during the first year and were anamnestic to the prenatal Questions 1–3. The Rs for these three scales were .02, .50, and .81, with $R_{av.}$ = .51. The fact that the R of .02 for Scale 1 (attitude toward being pregnant) is so much lower than the others suggests that a factor such as anxiety might have influenced the mothers' reports to this question during the first year. (This suggestion is strengthened by the fact that, when asked the same question in the recent anam-

---

[7] The corresponding Rs based on the physical measures were .91, .77, .91, and .96 for Scales 11–14, respectively, with $R_{av.}$ = .91. The difference between these two sets of Rs is due to the greater precision of the original measures.

nestic interview, the R between these two reports, which covered an interval of about 8 years, was .60.)

2. Short interval: sixth year to anamnestic interview. Scales 17–27 are based on questions regarding parental reports made when the children were 6 years old. The Rs for these scales range from $-.09$ to .53, with $R_{\text{av.}} = .24$. From Table 1 it is clear that this group of scales contains the lowest Rs, which finding we would not expect on the basis of the time factor alone.

There are at least two possible types of explanation for the generally low and variable Rs in this group. First, as we have already indicated, the interviewing and scoring procedures for the sixth year scales were not always strictly comparable for the historical and anamnestic scales. A second possibility is that the sixth year scales involved attitudes and criteria for the child's behavior which were themselves in a state of flux and perhaps were not distant enough in time to become stabilized. Thus, it may have been difficult for the mothers to differentiate their norms for their children at age 6 and at the time of the follow-up interview when they were from 7 to 8 years old. From this point of view, one might reason that the mothers were caught up in gradual changes in (or current uncertainties or conflicts about) their norms with respect to, for example, the extent to which the child should determine his own behavior (Scale 24) or choose his own playmates (Scale 27).

3. Long interval: prenatal to anamnestic interview. The Rs for Scales 1–10 range from .09 to .83, with $R_{\text{av.}} = .49$. This $R_{\text{av.}}$ would have been somewhat higher (.55) if the two anxiety scales (6 and 7) were excluded. Even so, the $R_{\text{av.}}$ of .49 for this 8-year interval is appreciably higher than the above $R_{\text{av.}}$ of .24 for the more recent, sixth year scales.

4. Long interval: first year to anamnestic interview. The Rs for Scales 11–16 range from .46 to .89, with $R_{\text{av.}} = .73$. This group includes, of course, the four "validity" scales; otherwise the Rs of .56 and .46 for Scales 15 and 16 approximate the $R_{\text{av.}}$ of .49 for the above group of long-interval scales.

One may speculate as to why, generally speaking, the coefficients indicating the relative consistency or agreement between the mothers' reports on two different occasions tend to be higher for the longer time intervals. But speculations aside, an inspection of Table 1 suggests that the type of scale involved played a more important role in determining the size of the coefficients than did the length of the time interval.

Types of Scales

Because of uncertainties regarding the comparability of the historical and anamnestic scales for the sixth year period, we will restrict our discussion of the various types of scales to those for which such questions do not arise, namely, Scales 1–16. These scales can be grouped under five headings, as follows:

1. "Hard facts." In our earlier discussion of the question of validity, we saw that the $R_{\text{av.}}$ for Scales 11–14 was .80, and that the coefficients for this group were higher than for any other grouping in spite of the rather long time interval. Perhaps the mothers were able to recall these facts so accurately because they were quantitative and relatively independent of their own attitudes, wishes, or anxieties.

2. General wishes and attitudes. Eight scales (2, 3, 4, 8, 9, 10, 15, and 16) had to do with the parents' wishes and attitudes, as to whether the child was wanted and planned for, the desired sex of the child, and, generally speaking, their attitudes toward various aspects of child rearing. The Rs for these eight scales ranged from .23 to .73, with $R_{\text{av.}} = .50$.

3. Anxiety scales. We may assume that during the prenatal period all the mothers were more or less anxious about their own and their child's well-being during the childbirth and whether their child would be healthy and undeformed. The Rs for Scales 6 and 7 were .09 and .32, with $R_{\text{av}} = .21$. Compared with the above two groupings, the Rs for these scales suggest that anxiety interfered with the mothers' accurate recall of their earlier attitudes regarding these matters.

4. Anxiety ratings on attitude scales. In rating the early historical material the statements of each mother relevant to each scale were also coded in terms of whether they touched upon areas or experiences that were markedly disturbing to her. On three scales (1, 4, and 7) there were 8, 6, and 4 of the 19 mothers, respectively, who were judged to show clear signs of anxiety. The Rs for these subgroups were .03 on Scale 1, .08 on Scale 4, and

$-.37$ on Scale 7, with an $R_{\text{av.}}$ of $-.09$ for the three scales. The corresponding *R*s for the mothers who were judged not to have been very anxious in connection with these matters were .17 on Scale 1, .56 for Scale 4, and .37 for Scale 7, with an $R_{\text{av.}}$ of .38.[8] The two sets of *R*s are in line with our previous findings that anxiety interferes with the accurate recall of this type of material.

5. Anxiety ratings on a nonattitude scale. The only other scale from the early material on which a sufficient number of mothers were judged to have shown clear signs of anxiety was Scale 5. The five mothers who expressed anxiety at the time about their difficulties during pregnancy did tend to recall their difficulties during the anamnestic interview ($R = .32$), whereas the nonanxious mothers tended not to do so ($R = -.37$). The apparent differential effect of anxiety on the reliability of recall of attitudes and of experiences will be touched upon again when we consider whether the mothers were having their first child or had had one previously.

Birth Order and Sex Differences among Children

As we observed earlier, the sample of families was selected for study before the birth of the children who also were to be studied over a period of several years. Under such conditions if, for practical reasons, one may not be able to reject the families initially selected, one runs the risk of getting a sample of children that is not balanced between boys and girls. Unfortunately, in the present study the birth order and sex of the children are confounded: first-born tends to be synonymous with girls and second-born tends to be synonymous with boys. Specifically, 8 of the 12 first-born children are girls and 6 of the 7 second-born children are boys.

Birth order of child. In discussing our findings with respect to birth order differences, we will limit our attention to two prenatal scales (5 and 6), the responses to which should not have been influenced by the sex of the unborn child, and to the 11 sixth year scales involving recent socialization attitudes.

On Scale 5, the 12 mothers who had their first child recalled their difficulties during pregnancy more accurately ($R = .57$) than the 7 mothers who had their second child ($R = .02$). On the other hand, on Scale 6 the mothers who had their first child tended not to recall their earlier reported anxiety over the childbirth process ($R = -.31$), as opposed to the mothers who had their second child ($R = .42$). These findings recall our previous discussion of Scales 5 and 6, where we cited data suggesting that anxiety served to facilitate accurate recall when associated with the mothers' experiences, but interfered with such recall when associated with earlier reported attitudes. One cannot be certain that mothers about to have their first child were more "anxious" than the others; they may have been only more involved and attentive to their pregnancy than the mothers who had been through the experience successfully once before. But whether it was anxiety or attentiveness, it appears that such factors influence differentially the recall of prior experiences and attitudes.

There were also differences in the reliability coefficients of mothers having their first or second child on the sixth year scales. The 12 mothers in the first group had an $R_{\text{av.}}$ of .10 for these 11 scales, whereas the corresponding $R_{\text{av.}}$ for the second group was .29. Furthermore, the individual *R*s were lower on 9 of the 11 scales for the mothers who had their first child. These findings are in line with the suggestion that the mothers' possible confusion or uncertainty about currently changing norms might interfere with the accurate recall of their previous attitudes in connection with this set of scales.

Sex of child. One may inquire as to whether mothers tended to give more reliable reports during the anamnestic interview, depending on whether their child was a boy or a girl—that is, whether it was of the opposite or the same sex as her own. Reasons for such an inquiry lie in the possible existence of culturally held attitudes that favor children of one sex more than the other and/or of less consciously held attitudes that favor children of the same or the opposite sex. But, because birth order and sex are confounded in this sample, we must consider any sex differences within the first- and second-born groups.

The relevant findings are summarized in

---

[8] The *R*s for all 19 mothers taken together were .60, .83, and .32 for Scales 1, 4, and 7, respectively, with an $R_{\text{av.}}$ of .63. The lower *R*s for both the anxious and nonanxious mothers is due to the fact that these two subgroups tended to be distributed at different ends of the six-point scales, so that *R* is raised when the subgroups are combined.

TABLE 2

AVERAGE RELIABILITIES FOR MOTHERS OF FIRST- AND SECOND-BORN BOYS AND GIRLS

Scales	First-born		Second-born	
	Boy ($n = 4$)	Girl ($n = 8$)	Boy ($n = 6$)	Girl ($n = 1$)
Prenatal (Scales 1–10)	.36	.53	.47	(.68)
First year (Scales 11–16)	.57	.71	.80	(.16)
Sixth year (Scales 17–27)	.02	.18	.03	(.45)
Total interview (Scales 1–27)	.32	.46	.44	(.54)

Table 2. Here we find that, for the first child, the mothers of girls tended to give more reliable reports during the anamnestic interview than did the mothers of boys. On the second child, however, the coefficients for the mothers of boys are about the same as those for mothers of first-born girls. But because of the small number of cases in each of these subgroups and other complicating factors,[9] these findings must be thought of as suggesting that birth order and sex differences might influence the recall of prior events in an anamnestic interview.

Individual Differences among Mothers

Finally, we can inquire about possible differences among the mothers in the extent to which their reports during the anamnestic interview agreed with what they had said earlier. In this connection, reliability coefficients indicating the agreement in the mothers' reports on the 27 historical and anamnestic scales were computed. These Rs range from .04 to .67, with $R_{\text{av.}} = .47$. Thus, although some mothers gave anamnestic reports that agreed reasonably well with what they had said previously, none of the mothers showed the degree of accuracy, or reliability, that one would hope for. Nor can it be overlooked that the anamnestic reports of some mothers were essentially random, at least within the limits of our scales. Furthermore, we find even greater variation among the mothers when we compute such reliability coefficients for the scales in each of the three time periods. Here we find that the Rs ranged from $-.39$ to .91 for the 10 prenatal scales, from $-.14$ to 1.00 for the 6 first year scales, and from $-.42$ to .55 for the 11 sixth year scales. Even allowing for the fact that these latter Rs may be relatively unstable because they are based on a small number of scales, one is nonetheless struck with the general *in*accuracy of the anamnestic reports given by many of the mothers.

The frequent low (or negative) coefficients and the wide variation among them raises the following question: Did the mothers tend to be consistently accurate or inaccurate in their reports involving different groups of scales, such as those in the three time periods? An answer to this question was arrived at by converting the 57 Rs of the 19 mothers for each of the three time periods to z scores (to obtain normalized scores) and then computing an R for the 19 mothers over the three time periods taken together. This R was $-.38$, which clearly indicates the mothers' inconsistency in the reliability of their reports over these periods. In other words, if the mothers tended to recall the scales for one time period more accurately than the average, they tended to recall the scales for the other time periods less accurately. This tendency is also seen when we correlate the above z scores for each pair of time periods. Here we find that the R is $-.07$ between the prenatal and first year periods, is $-.54$ between the first year and sixth year periods, and is $-.23$ between the prenatal and sixth year periods.[10]

Since the implications of these findings are all too obvious, we were led to inquire whether an interviewer could tell ahead of time which informants would be the "reliable" reporters of which class of prior events in an anamnestic interview. We first asked two persons who had interviewed the mothers and also were familiar with the project records to see if they could formulate on the basis of their subjective impressions any clues which might be helpful in this respect. Using this approach, they were not able to differentiate among the mothers in terms of the reliability coefficients for any of the time periods or for all 27 scales taken together.

---

[9] These subgroups also differ in several other ways, the relevance of which is not always clear for this study. For example, the parents of second-born children were more secure in their treatment of the child, the parents of girls were less well-adjusted to each other, even before the child was born; the mothers breast-fed their boys about twice as long as the girls and also tended to treat them better in various other ways as well. These and other birth order and sex differences in this sample are discussed in detail in a paper by Eckhoff, Gauslaa, and Baldwin (in press).

[10] The disconcerting finding that informants in an anamnestic interview tend not to give equally accurate reports for all scales has also been reported by other investigators using larger samples of subjects. In one study (Pyles, Stoltz, & Macfarlane, 1935), which was based on the reports of 252 mothers, the intercorrelations (r's) ranged from $-.18$ to .27 among five scales: month walked alone, month of first tooth, birthweight, weight at 12 months, and duration of labor. (We would have called the first four of the above "hard fact" scales; the authors report the highest r's—of .22 and .27—among the first three of the above scales. Also, the scales used in the present study cover a broader range than those cited above, which fact may account for the greater variation in the reliability coefficients.)

As a second approach, and on the basis of our other findings, we inquired whether the factor of anxiety in either its general or specific aspects might be related to the differences among the mothers in the accuracy of their recall. Global anxiety ratings by the sixth year and anamnestic interviewers were not significantly related to the above *R*s, nor were the scores on the anxiety scales. In the latter case, we combined Scales 6 and 7 from the prenatal period (anxiety over the child's appearance and childbirth) and Scale 17 from the sixth year period (security—or its converse, anxiety—in rearing the child). These scales also did not differentiate significantly among the mothers with high and low *R*s for any of the time periods or for the total set of anamnestic scales. However, the fact that we were not able by these procedures to find indices which are related to the mothers' reliability coefficients does not mean that such indices do not exist.

DISCUSSION

The findings of this study indicate that, in many respects, the mothers' statements during the anamnestic interview were not particularly accurate as reports of prior events. It appears that the anamnestic material did not reflect the mothers' earlier experiences and attitudes so much as their current picture of the past. This seems obvious enough. What did not seem obvious when we began this study is that the "reliability" of their picture of the past varied so greatly for different types of anamnestic material and from one mother to another. These findings do not, however, alter the fact that anamnestic interviews are usually conducted by persons who want to learn about what actually happened in the past.

An additional complication is the evidence that anxiety serves to distort the accurate recall of many experiences. In coming to a school or other guidance clinic, for example, the informant usually is anxious and unable to cope with some kind of problem, and so asks for help. In situations of this sort the interviewer inquires about the past so that he can make intelligent decisions about what kind of help should be given. It is an ironic possibility that in those situations where accurate information is most needed, it may be most distorted.

In this connection it is worth noting that the anamnestic interview, or some variation of it, has been used extensively to build up a body of "data" regarding, for example, the genetic roots of adult character structure. Since such "data" are frequently obtained from adult patients in psychotherapy, one may wonder about the extent to which they give an accurate, representative picture of the genesis of adult character structure in general. Perhaps the fact that many current theories of personality have been based on the reports of disturbed persons in therapy accounts in part at least for their predominant emphasis on conflicts and maladaptive mechanisms.

In any case, since the anamnestic interview is such an important practical and theoretical tool, its strengths and limitations should be investigated more thoroughly than we have been able to do. Even though the present study is ancillary to a broader project which was designed to meet other goals as well, we have been able to investigate the effects of certain conditions or sources of variation in anamnestic reports. On the basis of our findings it seems clear that more adequate investigations of the anamnestic interview can and should be carried out. To be most fruitful such studies should be designed with certain sampling, interview technique, and data analysis considerations in mind ahead of time.

Sampling. Because of the wide variety of situations in which the anamnestic interview is used, it would be desirable to use an experimental design with representative sampling of different types of informants, contents (attitudes, experiences, facts), etc. in different settings. Where possible, additional information should be collected to investigate, first, the "validity" as well as the "reliability" of anamnestic reports and, second, to identify factors that may be related to the marked individual differences among informants in the extent to which they recount accurately certain types of information.

Interview technique. In this study we have only alluded to problems of interview technique. Particular care should be taken in eliciting anamnestic material since what the informant says depends in part on the relationship between him and the interviewer and by how the questions are phrased by the one and interpreted by the other. In this connection, it is possible that some of our reliability coefficients would have been substantially higher if the project as a whole had not also attempted to achieve a variety of other goals. For example, while collecting the historical material, the mothers were sometimes permitted to digress from the questions to discuss related matters of concern to them. Although this procedure served other purposes of the project, it complicated this particular study, and was one of the reasons why the historical and anamnestic materials from the sixth year period are not strictly comparable and hence could not be coded with satisfactory precision.

Data analysis. Although indices such as reliability coefficients give a general idea of how this or that variable contributes to or interferes with the accuracy of anamnestic reports, this approach is relatively limited in the amount of information it provides. Given the proper experimental design, sampling procedures, and scaling techniques, one could approach the problem in terms of the major sources of variation (e.g., components of variance) and their interactions in anamnestic reports. This type of information would contribute more effectively to a better understanding of memory processes and also would enable the interviewer to estimate the relative accuracy of various types of anamnestic material obtained under specified conditions.

SUMMARY

This research was designed to study the effects of some of the factors which may influence the reliability of anamnestic reports. From a longitudinal study, records of 19 mothers' interviews before the children were born, when they were 1 and 6 years old, and other records were used to provide 27 normalized six-point scales. The scales were selected to cover relatively short (1 to 2 year) and long (7 to 8 year) time intervals and several different types of information. When the children were between 7 and 8 years old, the mothers were given an anamnestic interview which was built around the same 27 scales. The coefficient of intraclass correlation was used to give an index of the agreement or reliability of the mothers' reports between the earlier and the anamnestic interviews.

In general terms, we found that the mothers' statements during the anamnestic interview were not particularly accurate as reports of prior events. It appears that the anamnestic material did not reflect their earlier experiences and attitudes so much as their current picture of the past.

In more specific terms, we found that:

The length of time interval, by itself, was not meaningfully related to the reliability of the anamnestic reports.

Clear differences occurred from one type of information to another. The "hard fact" data, such as the length of child at birth were recalled best of all; general wishes and attitudes (not involving anxiety) were recalled next best; and attitude scales, concerning which of the mothers were anxious at the time, were recalled least well, with reliabilities around zero. If anxiety was associated with actual experiences, however, it seemed to facilitate the accurate recall of those experiences.

There were large individual differences among the mothers in the reliability of their anamnestic reports; furthermore, they were not consistent in the accuracy of their reports from one group of scales to another.

REFERENCES

CHAMPNEY, H. The measurement of parent behavior. *Child Develpm.*, 1941, **12**, 131–166.

ECKHOFF, EVA, GAUSLAA, J., & BALDWIN, A. L. Parental behavior toward boys and girls of preschool age. *Acta psychol., Amsterdam*, in press.

HAGGARD, E. A. *Intraclass correlation and the analysis of variance.* New York: Holt-Dryden, 1958.

PYLES, M. K., STOLTZ, H. R., & MACFARLANE, J. W. The accuracy of mothers' reports on birth and development data. *Child Develpm.*, 1935, 6, 165–176.

(Early publication received June 15, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 319-322

ACQUIESCENT RESPONSE SET AND THE PSYCHASTHENIA SCALE:

AN ANALYSIS VIA THE *AUSSAGE* EXPERIMENT

JAMES L. KUETHE
Johns Hopkins University

"I object. Counsel is attempting to lead the witness." This cry is familiar to everyone as part of the drama of the courtroom. Judges and lawyers have long been aware that the testimony of a witness may be biased if "leading questions" are permitted. Witnesses do not usually intend to give false information to a leading question but rather seem to be unaware that their testimony has been influenced by the examiner.

There have been two main approaches to the study of errors of testimony, the investigation of errors made during free recall and the investigation of errors induced by an examiner. A good review of early research on the topic is that of Whipple (1909). The *Aussage* demonstration has often been used in the classroom as a dramatic means of illustrating the low reliability of testimony. The typical procedure is for the instructor to have a confederate come up in front of the class and behave in some unusual manner. The class is then questioned about the events that have just taken place. The variability of the reports received convinces the class that testimony is subject to considerable error. This interesting pedagogical device only touches on one aspect of the problem of error of testimony.

The form of a question plays a major role in introducing error tendencies in the response (Burtt & Gaskill, 1932; Muscio, 1916), hence the legal concern aroused by "leading questions." However, there are considerable individual differences in susceptibility to influence of this nature that are ignored in studies interested only in showing that free recall is not reliable. The testimony of some persons is far more reliable than that of others. It was the aim of the present investigation to examine personality variables associated with individual differences in the accuracy of testimony.

Acquiescence to suggestion would seem to be involved when report is influenced by the form of a question. It is also obvious that most persons would not be as dramatically influenced if they were aware that pressures were being exerted upon them. Thus two personality variables appear to be important determinants of the extent to which a person's testimony may be influenced: an acquiescent response set and a tendency to be unaware of influence.

Kuethe (1959) has shown that persons with a strong acquiescent response set tend to cooperate with the experimenter and follow his instructions to perform a long uninteresting task. The acquiescent response set in this way relates to a tendency to comply with the dictates of another individual.

Eriksen, Kuethe, and Sullivan (1958) found that awareness of influence in verbal learning is related to scores on the Psychasthenia (*Pt*) scale of the MMPI. Low scorers on the *Pt* scale tended to lack verbal awareness of influence in a manner consistent with clinical theory and observation.

It was the hypothesis of the present investigation that persons both strong in acquiescent response set and low in score on the *Pt* scale would be more subject to influence when giving testimony than individuals not showing this personality pattern.

METHOD

The subjects were 230 students enrolled in an introductory course at Teachers College, Columbia University. During a class meeting all subjects were given the *Pt* scale of the MMPI and a measure of the acquiescent response set. The measure of the acquiescent or positive response set was essentially a measure of the tendency to agree to suggested concepts and is described in detail by Kuethe (1959).

Several weeks later a classroom "drama" was staged to reveal individual differences in fidelity of testimony. The sequence of events was as follows: the experimenter announced to the class that someone had not picked up her reports at the last class session. He then said: "Will Miss Sandra Moore please come up and get her report." The young lady, a confederate[1] of the experimenter came up on the stage and stated in a loud, irritated tone of voice that one of her reports was missing. When the experimenter stated that she must not have handed it in, she angrily shouted: "I have

---

[1] The author wishes to express his appreciation to Christina B. Gidynski for her assistance in the classroom "drama."

never heard anything so ridiculous." (At this outburst, the class gasped and directed their full attention to the stage.) Then, "Sandra" pulled a water pistol from the pocket of her skirt, shot the experimenter with a stream of water, and fled from the stage dropping a piece of paper as she went.

Five minutes later when the class was restored to order, they were questioned about the incident. They wrote answers to the following questions:

What was the young lady's name?
What color was her blouse? Her skirt?
What color was her purse?
Did she drop anything?

Actually there was no purse. In asking about it we have used an "implicative question," one that assumes or suggests the presence of an element not actually present.

RESULTS

All subjects were classified relative to the medians of the distributions of *Pt* scores and acquiescence (A) scores. Four classifications resulted: (*Pt*−, A+), (*Pt*−, A−), (*Pt*+, A−), and (*Pt*+, A+). For example, (*Pt*+, A−) would be the group of subjects above the median on the distribution of *Pt* scores and below the median on the distribution of A scores. The correlation between the *Pt* scores and the A scores in the present study was .34, $p < .01$ with 228 *df*. However, the variability of the scores was considerable with the result that the number of subjects in each group was almost equal.

The testimony of the subjects was scored for accuracy of report. The critical question was the implicative question. "What color was her purse?" This leading question influenced 34 subjects to indicate that there was a purse when in fact there was none. The remaining 196 subjects either left this question blank or wrote "none." The 34 influenced subjects did not differ from the other subjects in accuracy of response to the straight recall questions.

The central analysis for the present investigation was based on the distribution of the 34 influenced subjects in the two-variable personality classification. The frequency distribution shown in Table 1 produces a chi square of 11.05, $p < .02$ with 3 *df*. This result supports the hypothesis that persons both strong in acquiescent response set and low in score on the *Pt* scale are more subject to influence when giving testimony than individuals not showing this personality pattern.

TABLE 1

FREQUENCY WITH WHICH "INFLUENCED" AND "NONINFLUENCED" SUBJECTS BELONGED TO THE FOUR PERSONALITY PATTERNS

	(*Pt*−, A−)	(*Pt*−, A+)	(*Pt*+, A−)	(*Pt*+, A+)
"Influenced"	7	16	6	5
"Noninfluenced"	51	41	51	53

DISCUSSION

A person usually would not be "led" in his testimony were he aware that an attempt was being made to influence him. A tendency to be acquiescent should also contribute to susceptibility to influence of this comparatively subtle type.

There is considerable evidence that scores on the *Pt* scale are related to the ability to demonstrate verbal awareness. Scores on the *Pt* scale have been shown to be inversely related to "repression" (Eriksen & Davids, 1955; Kuethe & Hulse, 1960). After verbal conditioning, high scorers on the *Pt* scale are more able to verbalize awareness of influence than are low scorers as was shown by Eriksen, Kuethe, and Sullivan (1958). The tendency of the high scorer on the *Pt* scale to consciously monitor his behavior has been discussed by Kuethe (1961).

Acquiescent response set might logically be expected to be related to the *Pt* scale as the latter contains mainly positively keyed items. This is, however, not the case as has been shown by De Soto and Kuethe (1959). The clear implication of the *Pt* items that "something is wrong" (this scale contains "obvious" rather than "subtle" items) overrides any tendency to agree with an item because of acquiescence. Also it should be noted that the persons of primary concern in this investigation are below the median on *Pt* and above the median on A. These subjects would of course not be found were there a perfect positive correlation between A and *Pt*. Such a situation is consistent with the view that scores on the *Pt* and similar scales are a disjunctive function of symptoms (real or imaginary) and the capability of claiming these symptoms on an inventory (De Soto & Kuethe, 1960).

The less than perfect correlation between any performance measure and scales like the *Pt* scale and the Manifest Anxiety Scale is not entirely random error but is in part due to the fact that such scales are not unidimensional. When a scatter plot of *Pt* scores versus performance is examined, a triangular distribution is often noted (De Soto & Kuethe, 1960). The performance of subjects with low *Pt* scores is more variable than the performance of subjects with high *Pt* scores. This is to be expected if it is true that high scorers believe they have symptoms and have claimed them while low scorers may either lack the symptoms or have them but be unwilling or unable to claim them because of their defenses. High scorers are simply a more homogeneous group as far as personality dynamics are concerned.

It may well be that the A measure serves to identify those low scorers on the *Pt* scale who would have earned high scores had they not been prevented (consciously or unconsciously) by the implications of agreeing to the items on the scale. These subjects are using defenses related to repression or denial and as a result are not conscious of pressures from the environment and hence are likely to be influenced by a leading question. It is these subjects who do less conscious monitoring of their behavior and as a result show less verbal awareness of what has happened.

In contrast, those low scorers on the *Pt* scale who do not show acquiescent response set may be "true" low *Pt* subjects in the sense that while they are agreeing to all of the symptoms that they actually possess, they earn a low *Pt* score because they do not have many symptoms. These nondefensive subjects are more likely to be aware of influences from the environment and as a result to evaluate them critically.

Similarly, the high scorer on the *Pt* scale shows by the fact that he is capable of claiming symptoms on such an inventory that he is not using defenses of the repression or denial variety. He tends to consciously monitor his behavior and is perhaps, because of his higher anxiety, more likely to do so than are others. The tendency to monitor and maintain vigilance results in awareness of pressures from the environment which gives some immunity from influences such as leading questions and learning of the variety that requires that the individual not be aware that some aspect of his behavior is being reinforced.

A very critical point is the relation between *Pt* and acquiescence. While the high scorer on the *Pt* scale does show acquiescence, he does not earn his high *Pt* score because of it, in spite of the fact that most of the *Pt* items are keyed in the positive direction. The reliability of the *Pt* scale is not reduced when a semireversed scale is used. De Soto and Kuethe (1959) state:

> The intercorrelations between the oppositely keyed halves of the balanced *Pt* scale are so high, in fact, that acquiescence set can be ruled out as an important factor in *S*s' performance on the scale.

When *Pt* scores are correlated with scores on other positively keyed scales a positive correlation is found (Couch & Keniston, 1960). It is easy to assume, as did Couch and Keniston, that the correlations exist because acquiescence contributed to scores on all the scales. This is not necessarily the case. High scorers on the *Pt* scale earn high scores on positively keyed scales because acquiescence is part of their personality pattern, not because the items on the *Pt* scale are also positively keyed. High scorers on the *Pt* scale will deny the absence of symptoms just as readily as they will agree to the presence of symptoms; their behavior is essentially "symptom claiming" (De Soto & Kuethe, 1959).

There is a distinct possibility that the acquiescent response set may function in the way that the *K* scale of the MMPI was intended to function but does not, at least with respect to the *Pt* scale. The *K* scale cannot effectively "correct" the *Pt* scale for defensiveness. Fordyce (1956) reports a correlation of −.75 between the *K* and the *Pt* scales. The reliability of the *K* scale is .76 (Hathaway & McKinley, 1951); the *K* scale could not, therefore, show a closer relation to the *Pt* scale than it does. These two scales may be regarded as measures of the same thing, with the *Pt* scale being the better measure because it is more reliable. Study of the item content of these two scales reveals that many of the items refer to similar behavior and symptoms except that they are keyed in opposite directions, therefore resulting in the negative correlation. On both the *Pt* scale and the *K* scale a person has the same task: accepting or rejecting items that inquire about personal behavior and feelings. This view is supported by the correlations reported by Edwards (1957); the correlation between the Social Desirability Scale (SDS) and the *K* scale is .81 while the SDS and the *Pt* scale correlate −.85. As the split-half reliability of the SDS is .83, once again the picture is that of scales intercorrelating to the limits of their reliabilities.

In the present study, acquiescent response set has demonstrated its usefulness in identifying two distinct behavior patterns among low scorers on the *Pt* scale. This result supports the possibility of "suppressor variable" functions for the acquiescent response set, which has earned the right to be regarded as a personality variable and a very interesting one at that. Among its virtues it can be easily measured and with high reliability (e.g., a split-half *r*, without correction for test length, of .70 was reported by Kuethe, 1959. However, the present analysis indicates that it would be wrong to go to the other extreme and regard acquiescence as that which is being measured on most of the standard personality inventories. There is far more to an individual's personality than acquiescence; the important problem is the exploration of the ways in which this variable interacts with other personality variables to produce behavior variations from individual to individual and from situation to situation. This is, obviously, an area in which great caution is required; the investigation of a variable that does influence so much of behavior encounters unique problems. As was shown in the foregoing discussion, it is possible for a person to earn a high score on a positively keyed personality inventory, show acquiescence in a variety of situations, and yet have earned the high score on the inventory for reasons other than acquiescence. It would be easy to overlook this possibility and assume that acquiescence produced the high score on the positively keyed inventory—a case of the classic logical error of misattributing the direction of cause and effect between two correlated variables. It appears in the present situation that a person does not get a high score on the *Pt* scale because he is acquiescent, but rather, he is acquiescent because this is part of the personality pattern of the individual who earns a high score on the *Pt* scale.

SUMMARY

The classical testimony experiment was used to explore the relation between the acquiescence response set and a personality measure, the Psychasthenia scale of the MMPI. It was hypothesized that subjects who could be "led" in their testimony would have a strong acquiescence response set and be low scorers on the Psychasthenia scale. This hypothesis was supported by the testimony of 230 subjects who had witnessed a classroom "drama."

Evidence was discussed that the high scorer on the *Pt* scale is "acquiescent" but does not earn his high score on the positively keyed *Pt* scale because of this acquiescence. A "suppressor variable" function for the acquiescence response set was suggested.

REFERENCES

BURTT, H. E., & GASKILL, H. V. Suggestibility and the form of a question. *J. appl. Psychol.*, 1932, **16**, 358–373.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

DE SOTO, C. B., & KUETHE, J. L. The set to claim undesirable symptoms in personality inventories. *J. consult. Psychol.*, 1959, **23**, 496–500.

DE SOTO, C. B., & KUETHE, J. L. On the relation between two variables. *Educ. psychol. Measmt.* 1960, **20**, 743–749.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.

ERIKSEN, C. W., KUETHE, J. L., & SULLIVAN, D. F. Some personality correlates of learning without verbal awareness. *J. Pers.*, 1958, **26**, 216–228.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* (Rev. ed.) New York: Psychological Corp., 1951.

KUETHE, J. L. The positive response set as related to task performance. *J. Pers.*, 1959, **27**, 89–94.

KUETHE, J. L. The interaction of personality and muscle tension in producing agreement on commonality of verbal associations. *J. abnorm. soc. Psychol.*, 1961, in press.

KUETHE, J. L., & HULSE, S. H. Pessimism as a determinant of the tendency to claim undesirable symptoms on personality inventories. *Psychol. Rep.*, 1960, **7**, 435–438.

MUSCIO, B. The influence of the form of a question. *Brit. J. Psychol.*, 1916, **8**, 351–389.

WHIPPLE, G. M. The observer as reporter: A survey of the "psychology of testimony." *Psychol. Bull.*, 1909, **6**, 153–170.

(Early publication received October 12, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 323–328

AFFECT-STIMULATING EFFECTS OF COLORS[1]

ROBERT J. DRECHSLER[2]
Yale University

THE present study tested two hypotheses: colored stimuli elicit more emotional responses than do gray stimuli; different colors have different quantitative effects in this regard, with red the most stimulating. Although these contentions underlie much of the writing and clinical practice involving the Rorschach (1942) test, they have rarely been closely and systematically studied.

The controversy surrounding the alleged affect-stimulating characteristics of chromatic stimuli may be found in articles by Schachtel (1943), Rapaport, Gill, and Schafer (1946), Rickers-Ovsiankina (1943), Klopfer and Kelley (1942), Lazarus (1949), Siipola (1950), and Siipola, Kuhns, and Taylor (1952). Attempts to test the validity of the color-affect relationship have suffered from dependence on the Rorschach cards as crucial materials for testing the hypotheses. In very few cases (Haward, 1955; Holzberg, 1955) was the color variable removed from the specific context of the Rorschach cards. In every case, the criterion for emotional involvement of the subjects (*S*s) has been some measure of the way in which they solved problems *in which color played an integral part*. Such methods justify the criticism offered by Siipola (1950). As long as color is one element to be used in solving a problem, the differences observed between performances with colored versus achromatic stimuli may rest on differences in the difficulty of the two problems rather than on the emotionally stimulating properties of color. The present experiment attempted to overcome this methodological difficulty.

METHOD

The general technique was to show homogeneous rectangles of bright red, green, and gray on a screen, and to follow each presentation of a projected stimulus with an orally presented word-association test. The colors were presented under two conditions: continually through the word-association test, and in a 5-second exposure followed immediately by a word-association test. The use of long and short color exposures permitted evaluation of the disruptive effect colors may have while a task is being performed as compared with their continuing effect once they are removed.

For half the test exposures, the visual stimuli were exposed for 5 seconds, removed, and followed by an orally-presented five-item word-association list; for the other half the exposures, the visual stimuli were presented continuously through the oral presentation of a five-item list of words. In the latter instance, the exposure of each color began 5 seconds before the presentation of the first word in order to eliminate any disruption of response due to an abrupt presentation of the stimulus. The procedure also entailed a period of stimulation prior to word association that was identical with that for the noncontinuous color condition. One color (red, green, or gray) was exposed with each five-item word list. There were six stimulus conditions: red continuous exposure, red short exposure, green continuous exposure, green short exposure, gray continuous exposure, and gray short exposure. Each of the six conditions was presented four times to each *S*; in all, 20 words were presented with each stimulus condition, five at each presentation. There were, therefore, 24 five-item word lists, making a total of 120 words, which were taken from the Orbison (Rapaport et al., 1946, Vol. II, p. 13) and Kent-Rosanoff (Rosanoff, 1938) word lists for which relevant norms of responses are available (Lesser, 1948; Russell & Jenkins, 1954).

Each *S* served as his own control. The order of presentation of the six stimulus conditions was such that for each *S* each condition followed every other condition, except that no one color was ever shown on two consecutive presentations.

Between *S*s, the order of presentation was systematically varied, so that over the entire experiment, each of the six conditions occurred four times with each word list and in each sequential position. The 24 word lists were always presented in the same sequence; but within lists, the order of presentation of the individual words was systematically varied to control for possible effects due to position within the series.

A recall test for the entire list of 120 words was run for each *S* after all the association tests and stimulus conditions had been completed. Reaction times to the stimulus words for both the association and recall sections of the experiment were measured to .01 second with an electric timer. The timer was set in motion by the experimenter (*E*) when he finished pronouncing each word. It was stopped when he heard *S*'s voice.

*S*s were seen individually and were tested for red-green color blindness before the experiment. No *S* was color blind. The test room was dark except for a small shaded lamp on *E*'s desk. *S*s faced a screen from a distance of about 8 feet. The projected colored rectangles were 45″ × 32″. The word association instructions used by Rapaport et al. (1946), as adapted to the present purpose, were used.

---

[1] This paper is adapted from a dissertation submitted to Yale University in 1956 in partial fulfillment of the requirements for the PhD degree. The author wishes to express his deep appreciation to Seymour B. Sarason, Robert P. Abelson, and Roy Schafer for their many constructive suggestions in the design and execution of this investigation.

[2] Now at Pittsburgh Child Guidance Center.

All *S*s were male students in elementary psychology courses at Yale University. Selected at random from a list of students in the courses, they ranged in age from 18 to 23, the median and modal age being 19.

Results

In terms of the experimental design, two primary hypotheses were tested: there are more word-association disturbances under both red and green conditions than under gray conditions; and red conditions elicit more word-association disturbances than will green conditions.

In terms of relatively long reaction time (RT) to stimulus words as the index of disturbance, the results yielded highly significant differences ($p < .01$)[3] between colors. Responses under the red conditions had the longest RTs; responses under the gray conditions had the shortest. When the two degrees of freedom between colors are divided into Red-plus-Green versus Gray with 1 *df*, and the orthogonal comparison of Red versus Green with 1 *df*, both of these comparisons are significant at the .05 level of significance. The magnitude of differences between color groups was small but consistent. The average RT was 1.40″ for the red conditions, 1.31″ for the green conditions, and 1.27″ for the gray conditions. Of the 24 *S*s, 16 had their longest mean RT under red conditions, 5 under green, and 3 under gray. The chi square test (with Yates correction) of the deviation of this distribution from chance expectancy was significant at the .01 level. These results support Hypotheses 1 and 2.

Although over-all analysis of variance showed no significant differences between continuous and short exposures of colors, the data were further analyzed to determine whether the obtained results held equally for the continuous and short exposure times. The following additional analyses of variance were computed:

Performance RT—Continuous exposure only. Between colors was significant with p less than .01. For the comparison Red-plus-Green versus Gray, p fell between .05 and .10. For the comparison Red versus Green, p was less than .05.

Performance RT—5″ exposures only. Between colors was not significant, nor were Red-plus-Green versus Gray or Red versus Green. These results indicate that the obtained significant differences depend primarily on the long exposure conditions.

Relatively long reaction times in recalling the original word association responses. The analysis of variance between colors was significant with p less than .01. The comparison Red-plus-Green versus Gray was significant with p less than .01. Red versus Green was significant at the .05 level. The mean recall RTs were 1.32″ for red, 1.22″ for green, and 1.16″ for gray. Eighteen *S*s had their longest mean RT for red, 4 for green, and 2 for gray. A chi square test (with Yates correction) of deviation of this distribution from chance expectancy was significant at the .01 level. These results support the hypotheses.

As was the case with performance RTs, over-all analysis of variance showed no evidence of significant differences between continuous and short exposures of colors. The data were further analyzed as follows:

Recall RT—Continuous exposures only. Between colors was significant with p less than .01. Red-plus-Green versus Gray was significant with p less than .01. Red versus Green was not significant.

Recall RT—5″ exposures only. Between colors was not significant. Neither Red-plus-Green versus Gray nor Red versus Green was significant. These results suggest that the continuous exposure conditions contributed more than the short exposure conditions to the over-all difference between colors and to the distinction between the two chromatic colors and gray.

An over-all analysis of variance of incorrect recalls of initial responses, taken as a disturbance index, did not yield significant differences between colors. However, examination of the data indicated that although the means for the different color conditions were almost identical for the short exposure time conditions, the means for the continuous exposure time conditions showed systematic differences. Therefore, this latter half of the data was analyzed separately. Differences between means for the different color conditions were then found to be significant with p less than .05. Red-plus-Green versus Gray was not significant, although the differences were in the predicted direction. Red versus Green was significant with p less than .05. This finding supports Hypothesis 2 and is consistent with Hypothesis 1, but only for the continuous color exposure conditions.

Unusualness of response word, based on

---

[3] All p values are based on a two-tailed distribution.

norms from 1,000 university students for the Kent-Rosanoff word list (Russell, 1954) and on Lesser's (1948) norms from 400 New York City *S*s, not exclusively university students, for the Orbison word list, was also taken as an index of disturbance. For each stimulus word, the response words were listed in the order of frequency with which they were given in the normative group. With the responses arranged in descending order of popularity, cut-off points for scoring were established. Those responses which included 50% of the normative population (starting with the most popular response and adding frequencies down the list) were given an Unusualness score of 0; those responses which included the range from 50% to 75% of the normative population were given a score of 1; the least popular responses given by the remaining 25% of the population were scored as 2. Responses unique to the present experimental group were also given a score of 2. Analyses of variance were computed for the distribution of responses with weighted scores of 1 and 2 and for the Score 2–responses alone. These two analyses yielded practically identical results. Over-all analysis of variance showed no evidence of significant differences between color conditions. However, when the 2 degrees of freedom for the Between Colors measure were broken down into Red-plus-Green versus Gray and Red versus Green, the former measure was significant at the .05 level. These results support Hypothesis 1.

In addition, a score was devised based on the responses of the 24 *S*s in the experiment. Each response that occurred only once was considered "unique" and was given a weighted score. The weight was the reciprocal of the total number of different responses to the particular stimulus word in this group of *S*s. The weights represented an attempt to reflect arithmetically the idea that a unique response to a word that ordinarily yields but few different responses rates a higher "disturbance" score than a unique response to a stimulus word that elicits a wide variety of responses. Analysis of variance showed no significant differences between colors. Although the rank order of red, green, gray did not hold for this measure, red and green both elicited more unique responses than did gray. The results were not statistically significant, however.

Finally, unusual response words were rated as disturbed or not disturbed by a psychologist[4] who was supplied with a list of all responses to all stimulus words. Those responses which had received a 0 score for Unusualness were omitted from consideration. Also omitted were all responses which had a frequency of 25 or more on the Russell and Jenkins (1954) norms for the Kent-Rosanoff list, or a frequency of 10 or more on the Lesser (1948) norms for the Orbison list. The rater did not know the experimental condition under which a response was obtained, nor did he see the pattern of responses given by any individual *S*. Analysis of variance for this measure showed no significant differences between groups; hence no detailed analysis or reliability study of this qualitative measure was attempted. However this measure, in common with all other measures, showed gray conditions to elicit the fewest disturbed responses.

Analysis of variance for each of the six disturbance measures also tested differences between *S*s and the various interactions of colors, *S*s, and durations of exposure. For every disturbance measure, the differences between *S*s were highly significant with p less than .001. On the performance RT measure, the interaction of Colors × Subjects approached significance at the .05 level. This interaction, under continuous exposure conditions only, was significant at the .01 level. Different *S*s thus reacted differently to the different colors despite the fact that the preponderant tendency was for the rank order of disturbance scores on this measure to be Red, Green, and Gray. No other main comparison or interaction approached statistical significance on any of the disturbance measures.

After each disturbance measure had been analyzed, rank-order correlations of performance were computed, a rank being assigned to each *S* according to the magnitude of his contribution to the obtained differences between the color groups. In this way a measure was obtained of tendency for differences found on different measures to stem from the same *S*s. In each measure, the total score for each *S* was used to ascertain his rank. Table 1 shows that there was a significant tendency for the

---

[4] The writer thanks Roy Schafer for making these ratings.

TABLE 1
RANK-ORDER CORRELATIONS AMONG THE SIX TEST MEASURES

	RT(P)	RT(R)	IR(C)	IR(T)
RT(P)				
RT(R)	.49*			
IR(C)	.25	.26		
IR(T)	.23	.26	.55**	
UR	.12	.30	.24	.44*

Note.—Abbreviations:
RT(P)—Reaction time-performance
RT(R)—Reaction time-recall
IR(C)—Incorrect recall-continuous exposure only
IR(T)—Incorrect recall-total scores
UR—Unusualness of response words.
* $p < .05$, two-tailed test.
** $p < .01$, two-tailed test.

same *S*s to contribute to disturbance measures on both performance and recall RT and on unusualness of response and incorrect recall. The significant correlation between IR(C) and IR(T) is an artifact; IR(C) is contained within IR(T).

Further analysis of disturbance measures for each *S* was performed to determine whether the different types of disturbance shown by each *S* tended to cluster together in reaction to specific words and specific color conditions. On single stimulus words, scores for each *S* were dichotomized as scored plus or minus, depending on whether the response was "disturbed" or not. Chi square tests were performed between the measures for each *S* and summed for all *S*s. Summing the individual chi squares was justified by the fact that for all comparisons and all *S*s the predominant tendency was for both measures to be in the same direction, i.e., both disturbed or both not disturbed. For RT measures, dichotomization was accomplished by dividing the RTs at the median, with medians for each *S* computed separately for his red, green, and gray conditions. RTs that were longer than the median RT (for *S* on each color condition) were scored plus; those below the median were scored minus. For other measures, each response was scored "disturbed" or "not disturbed." Only measures that had shown significant results on analysis of variance were compared in this way. The measures paired in this way included RT (Performance) versus RT (Recall), RT (Performance) versus Incorrect Recall, RT (Recall) versus Incorrect Recall, RT (Performance) versus Unusualness of response, RT (Recall) versus Unusualness of response, and Unusualness of response versus Incorrect Recall. For every pair of measures, the summed chi squares were significant at the .01 level or better.

DISCUSSION

The results generally confirm Hypotheses 1 and 2. All the analyses give results consistent with Hypothesis 1; analyses of RT data for both performance and recall, as well as of the Incorrect Recall data for the continuous exposure condition, support Hypothesis 2. The most striking differences between color conditions were obtained when the visual stimuli were present before and during the time the word association lists were presented. This finding suggests that with a normal population there is rapid recovery from the emotionally stimulating effects of colors as soon as the colors are removed. The direct relation to the color variable of length of RT and occurrence of unusual responses on the performance section of the present experiment appears to be established.

A question may be raised, however, as to whether recall disturbances bear a similar direct relation to the color variable. The disturbance measures clearly were not independent of each other. There was a significant tendency for stimulus words to elicit responses which were rated "disturbed" on more than one measure, and this tendency was demonstrated for every pair of measures. During the recall part of the test the colored slides were not shown. Recall RT and incorrect recall of initial responses were both highly related, within *S*s, to performance RT. These data suggest that recall disturbances may have depended on the reactions occurring during the performance section rather than on a delayed effect of the colored stimuli themselves.

The significant correlation between *S*s of unusualness of response and incorrect recall indicates that an unusual response is harder to recall than a usual one. This fact alone would account for increased RT on recall for unusual words. *S* tries to remember the initial response, sometimes fails to do so, and finally gives an incorrect response. In this way the high correlation between recall RT and incorrect recall can be accounted for without recourse to the color variable. The crucial disturbance here is probably the unavailability of the original response; it is the test instructions that shift the emphasis to recall RT and incorrect recall.

All the disturbance measures used in this study have previously found some validation as emotional indicators. The validation has come primarily from comparisons of diagnostic groups with normals and of emotionally loaded stimulus words with neutral words (Hull & Lugoff, 1921; Jung, 1919; Luria, 1932; Rapaport et al., 1946; Rosanoff, 1938; Schumer, 1949). In the present experiment, these measures were used with a relatively homogeneous normal population and with a word list from which most of the "critical" words had been removed.[5] It is not surprising, therefore, that the magnitude of the differences between groups should be much less for the present experiment than for the groups on which validation of the word association test was based. The purpose of neutralizing the word list and of using normal *S*s was to minimize the variance in disturbance scores due to factors other than the color variable.

Rapaport et al. (1946) have elaborated the idea that evidence of selective disturbance in the word association test generally points to the stimulus word as the source of disturbance. In their findings, the specific response word usually did not reveal the nature of the disturbance, but only that some disturbance had occurred. In the present experiment, stimulus conditions must be looked to for the source of disturbance. The consistency with which disturbance differences occurred between color conditions is evidence for this.[6]

It may be argued that *E*, with full knowledge of the hypotheses, may unwittingly have influenced such a delicate measure as RT in the desired direction. That this was not the cause of the differences between color conditions is indicated by the following considerations:

1. RT was not the only measure to show systematic differences between colors.
2. Within-subject chi squares between RT and other disturbance measures were consistently significant, indicating that stimulus words which elicited long RTs also elicited other signs of disturbance.
3. Mean differences between RTs for the different colors were of the magnitude of .10 second, which would have required a most sensitive preconscious mechanism to produce.

Results of this experiment challenge Siipola's (1950) hypothesis that color itself has no emotional impact and that difficulty in handling Rorschach color cards is due only to the increased difficulty of the problem presented when color has to be integrated with form. In Siipola's formulation, emotional accompaniments of color responses stem from the *S*s' ego involvement in a difficult problem, not from intrinsic properties of colors. The present experiment did not require *S*s to integrate color into their responses. If the differences between colors and gray were due to *S*s' tacit assumption that the visual stimuli should be incorporated in their responses, it is unclear why chromatic stimuli should have presented more of a problem than gray because the visual stimuli were homogeneous and rectangular, and the stimulus words had no particular reference to red, green, or gray.

A possible explanation for the role of colors as affect-stimulating agents is that bright colors intrude on the perceiving person in such a way that his ongoing thought processes are interrupted or disrupted until the intrusive stimulus can be removed or otherwise assimilated. This conception of the intrusive nature of color has been elaborated by Schachtel (1943). The intrusion of color is similar to the effects of more specifically threatening environmental or internal stimuli, and for this reason it may either exaggerate strong defenses that have been developed against such threats or disrupt these defenses if they are weak.

The prepotent responses to different colors may themselves be disturbing in different degrees. For example, the common associations of "blood," "fire," "anger," and "passion" with red, and the disturbing connotations of these associations may account for the consistency with which red stimuli elicit the longest RTs. The significant interaction of Colors × Subjects under continuous exposure conditions for performance RT indicates that different *S*s react differently to different colors. This argues against the universal efficacy of one color over another or over gray, and it is consistent with the idea that individual associations to a particular color or individual differences in type and efficacy of defense measures may determine the extent to which that color produces measurable disturbances in word association responses.

---

[5] In selecting 120 words for which relevant normative data existed, it was impossible to eliminate all potentially critical words. Possibly any such word list, even without restrictions, would contain critical words. In the present list the words "love," "bite," "health," and "hungry" clearly pertain to common areas of conflict.

[6] A significant interaction of Colors × Stimulus words could account for the results, although words obviously related to any of the visual stimuli were not included on the list. This interaction was not tested, and in any case could not have invalidated the general findings that red elicited more disturbances than green, and both red and green elicited more disturbances than gray.

The conception of color disturbance presented here is not intended to contribute to the *mystique* surrounding some interpretations of Rorschach color shock. It is not suggested that color is the only variable that may produce the disturbances observed. Any sharply intrusive variable would be expected to have a similar effect.

SUMMARY

It was the purpose of this experiment to test the following hypotheses: colored stimuli elicit more emotional responses than do gray stimuli; different colors have different quantitative effects in this regard, with red being the most disturbing color. While these hypotheses are central to Rorschach test interpretation, they have not previously found clear-cut experimental support.

The method used was the projection on a screen of homogeneous rectangles of bright red, green, and gray. Each presentation of a visual stimulus was followed by an orally-presented word-association test. Several measures of disturbance on this test were used to measure disturbance elicited by the visual stimuli.

Several disturbance measures revealed statistically significant differences between colors and all supported Hypothesis 1. Reaction time measures showed significant differences in support of Hypothesis 2. The measures of disturbance proved to be interdependent, a fact which created some ambiguity with regard to details of interpretation, but did not interfere with the major conclusion that the hypotheses were supported.

The results were tentatively interpreted in terms of an "intrusion" hypothesis. Colors were thought to intrude on the perceiver in a manner similar to the intrusions of more specifically threatening environmental or internal stimuli and to arouse or break through the defenses used against threats. The degree of disturbance caused by a particular color was thought to depend on the disturbing connotations of the associations it arouses or on the type and efficacy of defense stimulated thereby.

REFERENCES

BECK, S. J. *Rorschach's test.* Vol. I. New York: Grune & Stratton, 1944.

HAWARD, L. R. C. Colour associations as a contributing factor to neurotic colour shock. *Brit. J. med. Psychol.*, 1955, **28**, 183–187.

HOLZBERG, J. D., & SCHLEIFER, M. J. An experimental test of the Rorschach assumption of the impact of color on the perceptual and associative processes. *J. proj. Tech.*, 1955, **19**, 130–137.

HULL, C. L., & LUGOFF, L. S. Complex signs in diagnostic free association. *J. exp. Psychol.*, 1921, **4**, 111–136.

JUNG, C. G. *Studies in word association.* New York: Moffat, Yard, 1919.

KLOPFER, B., & KELLEY, D. *The Rorschach technique.* Yonkers: World Book, 1942.

LAZARUS, R. S. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.*, 1949, **44**, 506–516.

LESSER, G. S. Commonality of response to the Orbison Word-Association Test. Unpublished master's thesis, Columbia University, 1948.

LURIA, A. R. *The nature of human conflicts or emotion, conflict and will.* (Trans. by W. H. Gantt) New York: Liveright, 1932.

RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.

RICKERS-OVSIANKINA, M. Some theoretical considerations regarding the Rorschach method. *Rorschach res. Exch.*, 1943, **7**, 41–53.

RORSCHACH, H. *Psychodiagnostics: A diagnostic study based on perception.* (Trans. by P. Lemkau & B. Kronenberg) New York: Grune & Stratton, 1942.

ROSANOFF, A. J. *Manual of psychiatry.* New York: Wiley, 1938.

RUSSELL, W. A., & JENKINS, J. J. *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test.* Minneapolis: Univer. Minnesota, 1954.

SCHACHTEL, E. G. On color and affect. *Psychiatry*, 1943, **6**, 393–409.

SCHUMER, F. C. Some behavioral correlates of Rorschach human movement response. Unpublished doctoral dissertation, Yale University, 1949.

SIIPOLA, ELSIE M. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.

SIIPOLA, ELSIE, KUHNS, F., & TAYLOR, V. Measurement of the individual's reactions to color ink blots. *J. Pers.*, 1952, **21**, 22–47.

(Received May 11, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 329-334

AUTHORITARIANISM AND MEN'S REACTIONS TO SEXUALITY AND AFFECTION IN WOMEN[1]

RALPH ROTHSTEIN
Michael Reese Hospital, Chicago, Illinois

THIS paper was stimulated by Gollin's studies (1951, 1954) on the formation of impressions of personality. In these experiments male subjects (*S*s) were shown a movie consisting of scenes in which the same woman portrayed two diverse roles, that of the prostitute and the kindly woman. Gollin found that impressions of personality were formed along three distinct lines. Some *S*s described the actress in terms of only one character quality and ignored the other (simplified or unitary impressions). Others characterized her in terms of both behavioral themes but did not relate them (aggregated impressions). There were still others who, in addition to noting the divergent behavioral patterns, integrated them in a meaningful and consistent manner (integrated or related impressions). Gollin has proposed that such differences in impression formation can be accounted for in terms of "cognitive-organizing" processes in the perceiver (Gollin, 1954, 1958; Gollin & Rosenberg, 1956) and has begun to explore the possible sources of these differences. In his recent experiment with children he found that age, sex, intelligence, and social background were all significant factors.

Authoritarianism should also be related to different cognitive modes of organizing impressions of others. If we assume that the problem of forming an impression of someone who displays inconsistent or disparate behavioral tendencies represents an ambiguous perceptual task, it follows that authoritarian persons, because of their presumed intolerance of ambiguity, should tend to form nonintegrated impressions. This formulation seems applicable regardless of the specific content of the judged person's inconsistencies. Gollin was apparently not concerned with the specific role of content in organizational patterns, and did not discuss why he selected promiscuity and kindness in a woman as the diverse character qualities, in preference to other discrepant characteristics, e.g., stinginess vs. generosity. However, in the present study, the specific behavioral characteristics (sexuality and kindness) displayed by the stimulus person were regarded as relevant, apart from the formal inconsistency of behavior, as groups differing in authoritarianism were expected to react differentially to the sexual and kindly cues provided by the stimulus person.

Among the variables found by Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) to be associated with authoritarianism was an exaggerated concern with sexuality, which the authors suggest, reflects a tendency to project unacceptable sexual impulses on to others. Frenkel-Brunswik's clinical interviews further showed male authoritarians to possess "dichotomous sex attitudes"; that is, they separated, or were not able to integrate, sexuality and affection in feeling and behavior.

Authoritarianism was therefore considered to be a personality characteristic of the perceiver significant to the process of social perception, especially when the stimulus person's behavior connotes sexuality and kindness (affection). While the importance of authoritarianism as a variable in impression formation has been demonstrated in several investigations (e.g., Jones, 1954; Scodel & Freeman, 1956; Scodel & Mussen, 1953; Thibaut & Riecken, 1955), these studies have dealt with aspects of behavior relating to leadership, insightfulness, aggression, etc., rather than with the role of authoritarianism in men's social perception of women.

On the basis of the findings emerging from *The Authoritarian Personality*, the following hypotheses were framed:

1. Authoritarians should tend to overemphasize the sexual aspects of a woman's behavior, even though she is also shown to be affectionate and kindly as well.

2. Authoritarians should show dichotomous

---

[1] This research is based in part on a dissertation submitted in partial fulfillment of the requirements for the PhD degree, University of Massachusetts, 1959.

The author wishes to thank Solis L. Kates, Seymour Epstein, and Claude C. Neet for their helpful criticisms and suggestions in carrying out the investigation.

sex attitudes as evidenced by their "splitting" of sexuality and affection. Further, their impressions of the personality of a woman who is both sexual and kindly should tend to be organized in a nonintegrated manner.

METHOD

Subjects

One hundred and twenty male undergraduates enrolled in the introductory psychology course at the University of Massachusetts were used as *S*s. Of this group 32 were selected as high authoritarians and 32 as low authoritarians.[2]

Stimulus Material

The stimulus situation consisted of a silent motion picture containing two scenes. One person, a young, attractive female in her middle twenties, appeared in both scenes and henceforth will be referred to as the star. In Scene 1 the star portrayed the "sexy" female, whose physical appearance and behavior were intended to characterize her as sexually appealing. Although the star was seen as sexually attractive, her behavior was not intended to suggest sexual immorality or promiscuity. In Scene 2, the star's behavior characterized her as being kind, affectionate, and understanding. An attempt was made to avoid any behavior that might be interpreted as sexually suggestive. It was intended that the star's behavior be equally convincing in both scenes, so that "objectively" one could describe her as both sexual and kindly to about the same degree. Each scene was approximately 5 minutes in length. The movie was presented to three other psychologists for judgment and there was agreement that the kindly cues were at a minimum in the sexual scene and sexual cues at a minimum in the kindly scene. A brief description of each scene is presented below:

Sexual Scene. As the scene opens, a young woman in a black slip is sprawled out on a bed. Her eyes are closed and her hair is somewhat disheveled. After a few seconds, she gets out of bed, yawns lazily and proceeds to groom and dress herself. Her mannerisms are sensually suggestive (e.g., she admiringly passes her hands over her body, carefully strokes her hair, etc.). She puts on her stockings, paying close attention to whether the seams are straight. She is shown attaching the stockings to her girdle, so that her thighs are fully revealed. At one point in this sequence, the camera is raised above her head and focuses on her breasts, which at this angle are partially exposed. She then puts on a tight fitting, low-cut dress and is shown carefully applying make-up. As the scene closes she is shown embracing her date, who has called for her; her general attitude is one of indifference.

Kindly Scene. As the scene opens, the star is reading a book. She is wearing a loose fitting jumper and blouse, so that her bodily attractiveness is minimized. On the table close to her is the photograph of a young man in army uniform. As she looks at it, a tender, affectionate expression comes over her face. After getting some stationery, she commences to write a letter, the content of which clearly reveals her love for the man in the photograph. The following is an excerpt from the letter which was, in its entirety, presented to *S*s:

> Although you never complain in any of your letters, I know that things are pretty rough out there. I only wish that there was something I could do to help to make you feel the warmth of home and to show you how much I love you. Please, darling, take care of yourself; you mean everything to me.

As she finishes the letter, a boy of about 13 enters the room and is greeted by the star, who messes his hair in an affectionate gesture. He sits down with a magazine and the star leaves the room, returning shortly with a snack for him. After eating, the boy begins his school work. He has difficulty with his lesson and goes to the star, who helps and encourages him with his work.

Procedure

The California F Scale (modified Form 45) was administered to all 120 *S*s. Of the 30 items on the F Scale employed in this study, 26 were the same as those in Form 45, with the other 4 items only slightly modified. The mean F score for the high and low authoritarian groups, each composed of 32 *S*s, was 130.25 and 101.63, respectively.[3] Two weeks after the F scale was administered, *S*s were shown a silent movie in groups of no more than 12. Two treatment conditions were used in presenting the movie. Half the group was shown the sexual scene first (SK presentation) and for the remaining half, the kindly scene was presented first (KS presentation). After the movie was shown, *S*s were asked to write their impressions of the personality of the woman who appeared in the movie.

Written Impressions

Gollin's (1951) criteria were used in categorizing the impressions into type of organization employed. In brief, these were: (*a*) unitary—in which one major theme (the star's sexuality or kindness) dominates the impression; (*b*) aggregated—in which the two major themes are present, but treated as independent and unrelated aspects; (*c*) integrated—in which both themes are present, each being related to the other, while the independence of each is maintained. To study the overall affective tone of the written impressions, each protocol was classified according to whether it was favorable, unfavorable, or both favorable and unfavorable. Organizational type and affective tone were judged independently of each other. In classifying the impressions, numbers were substituted for *S*s' names and all other identifying information was concealed.

---

[2] The 64 *S*s who participated in this study were originally selected on the basis of their performance on a specially constructed set of TAT cards depicting mother-child interactions. An attempt was made to differentiate *S*s on the basis of approach-avoidance conflict over sexual feelings toward their mothers. This measure, however, was only partially successful. *S*s were significantly differentiated only on the basis of the avoidance measure. The specific hypotheses which were then formulated for this "conflict" group were not substantiated. A complete discussion of this aspect of the research can be found in the author's doctoral dissertation, Men's Perceptions of Sexuality and Kindness in a Woman, University of Massachusetts, 1959.

[3] Only the F Scales of those 64 *S*s already selected to serve in the experiment on the basis of their performance on the TAT measure of "avoidance sexual conflict" toward their mothers were scored. In both authoritarian groups, there was an equal number of *S*s who were considered high and low in avoidance conflict over feelings toward mother.

Movie Rating Scale

To complement the free written description, a rating scale was administered after the written impressions of the star had been collected. The rating scale contained 30 randomly arranged items, 15 relating to sexual behavior, and 15 to kindly and affectionate behavior, each of which *S* was to mark according to one of four categories: strongly agree (+2), agree (+1), disagree (−1), and strongly disagree (−2). Examples follow, the first three sexual and the latter three kindly:[4]

If she were to buy a bathing suit, she would probably select one that would be revealing and tight fitting.

She has a great deal of flashy sex appeal.

The way she walks and dresses probably attracts men's interest in her.

She probably would enjoy doing charity work in which she could help others.

She appears to be a sweet, understanding young woman.

She appears to be the type of woman who, if she quarreled with her husband, would probably feel badly and want to make up.

Prior to its use in the study proper, the items of the scale were administered to 10 male graduate students who were asked to indicate whether an item indicated kindness or sexuality in a woman. Only the items with at least 90% agreement were retained. On all items but one, judges were unanimous.

Heterosexual Questionnaire

A questionnaire of 12 items indicating splitting of sexuality and affection in heterosexual behavior was constructed by the experimenter (*E*) and submitted to three other psychologists for judgment. The final form of the questionnaire consisted of the eight true-false statements that were considered most appropriate. These also allowed four responses: definitely true (+4); mostly true, but not entirely so (+3); mostly false, but not entirely so (+2); and definitely false (+1). The questionnaire was presented after the movie rating scale was completed, with strong assurances of confidentiality. The items were:

1. I believe that a girl who likes to look sexually attractive is unlikely to be a devoted wife when married.
2. I have found (or probably would find) it very difficult to fall in love with a girl with whom I have had intercourse.
3. From my experience, I would say that all women are fundamentally "sexual creatures" and that when they appear to be decent and kind it is only a façade.
4. My feeling that there is something basically animal-like and crude associated with sex interferes (or would interfere) with my expressing myself sexually toward someone I like very much.
5. I find myself losing my respect (or I would lose my respect) for a girl who lets me neck or pet with her on our first date.
6. From my experience I would say that in general very few sexy girls have sweet, decent dispositions.
7. I find that it is easy (or would find it easy) to get sexually aroused with a girl who has had sexual experience, but find it (or would find it) difficult to express my sexual feelings to a girl I love and respect.
8. I feel that basically there are two types of girls: the sexy type and the nice, respectable type.

RESULTS

Perception of Sexuality and Affection

To test the hypothesis that high authoritarians would perceive the star as more sexual than kindly, a mixed Type III analysis of variance (Lindquist, 1953) was employed. The two dependent variables were the scores for each *S* on the sexual and kindness subscales of the movie rating scale. Analysis indicated that regardless of group differences, *S*s as a whole did not perceive the star as being either significantly more sexual ($M = 43.47$) or more kindly ($M = 40.28$). Order of presentation of the movies also was found not to influence *S*s' movie ratings. Significant differences at the .025 level, however, were found in *S*s' sexual and kindness ratings when the effects of authoritarianism were evaluated ($F = 5.34$; $df = 1$). As can be seen in Table 1, high authoritarians rated the star as significantly more sexual than kindly, while the ratings of low author-

TABLE 1

COMPARISON OF SEXUAL AND KINDNESS SCORES FOR THE HIGH AND LOW AUTHORITARIAN GROUPS

Authoritarian	*N*	Sexual Scores		Kindness Score		*t*
		M	*SD*	*M*	*SD*	
High	32	45.72	6.39	38.69	9.45	3.41*
Low	32	41.22	6.53	41.88	6.29	.41

* Significant at beyond the .001 level for 62 *df*.

---

[4] A copy of the complete scale has been deposited with the American Documentation Institute.

itarians were practically identical on the two variables.[5] Further, high authoritarians perceived the star as significantly more sexual than did low authoritarians ($t = 2.73$; $df = 32$; $p = .01$). Inspection of Table 1 also reveals that low authoritarians tended to rate the star as more kindly than did high authoritarians, but this difference fell short of significance ($t = 1.56$; $df = 32$; $p = > .10$).

Dichotomization of Sexuality and Affection

Three measures were used to assess the splitting of sexuality and affection. Two of these, ratings of sexuality and kindness, and the organizational aspects of the written impressions, were based on the movies, while the third pertained to *S*s' self-reports of their heterosexual behavior and attitudes.

It was found that the groups high and low in authoritarianism did not differ significantly on the movie measure of splitting, expressed as the absolute difference between each *S*'s sexual and kindness ratings.

The written impressions of the star were classified into organizational type according to Gollin's (1951) criteria. Interjudge agreement based on each of two judges independently classifying 52 impressions was 78.8%. Results presented in Table 2 suggest that authoritarians tended to organize their impressions in a nonintegrated, that is aggregated or unitary manner, relative to the low authoritarians, the majority of whom formed their impressions in a more integrated style. However, the significance of this finding as determined by a 2 × 2 chi square analysis between authoritarianism and integration of impressions (integrated vs. nonintegrated categories) was > .10 ($\chi^2 = 1.82$; $df = 1$).

When the written impressions were analyzed for affective tone (see Table 3), there was again a tendency, though nonsignificant, for high authoritarians to characterize the star more unfavorably than favorably, relative to the low authoritarians. Interjudge agreement, based on 52 randomly selected protocols, independently classified by two scorers, was 80.8%.

TABLE 2

DISTRIBUTION OF IMPRESSION CATEGORIES FOR HIGH AND LOW AUTHORITARIANS

Authoritarianism	Organization of Impressions		
	Integrated	Nonintegrated	
		Aggregated	Unitary
High	7	15	10
Low	13	11	8
N	20	26	18

TABLE 3

CHI SQUARE ANALYSIS BETWEEN AUTHORITARIANISM AND FAVORABLENESS OF IMPRESSIONS

Authoritarianism	Favorableness of Impressions		
	Favorable	Unfavorable	Both Favorable and Unfavorable
High	8	13	11
Low	14	7	11
N	22	20	22

Note.—$\chi^2 = 3.44$; $df = 2$; $p = > .10$.

TABLE 4

EFFECTS OF AUTHORITARIANISM AND SEQUENCE PRESENTATION ON THE DEGREE OF "SPLITTING" OF SEXUALITY AND AFFECTION IN REPORTED HETEROSEXUAL BEHAVIOR

Source	*df*	*SS*	*MS*	*F*
Authoritarianism (F)	1	129.40	129.40	12.72**
Sequence Presentation (S)	1	40.65	40.65	3.997*
F × S	1	13.13	13.13	1.29
Within	60	610.06	10.17	
Total	63	793.24		

* $p = .05$ for $F = 4.000$.
** $p < .001$ level.

The findings for the effects of authoritarianism and sequence of presentation on degree of splitting of sexuality and affection in *S*s' reported heterosexual behavior and attitudes are presented in Tables 4 and 5. As can be seen, both authoritarianism and sequence of presentation were significantly related to degree of dichotomization in *S*s' perceived behavior.

---

[5] An *F* test was performed between the variances of the high authoritarians' sexual and kindness scores and heterogeneity of variance was indicated ($F = 2.19$; $df = 31, 31$; $p = .04$). One should regard the obtained *t* value as somewhat inflated. However, since the significance of *t* was beyond the .001 level, it seems safe to conclude that a genuine difference does exist between the means of the sexual and kindness scores.

TABLE 5

HETEROSEXUAL QUESTIONNAIRE SCORES IN RELATION TO AUTHORITARIANISM AND SEQUENCE OF PRESENTATION

Authoritarianism	Heterosexual Questionnaire Score		
	N	*M*	*SD*
High	32	18.03	3.47
Low	32	15.19	2.94
Sequence of Presentation			
SK	32	17.41	3.19
KS	32	15.81	3.64

High authoritarians showed a greater tendency to separate sexual impulses and feelings of affection in their reported heterosexual relationships, relative to low authoritarians. *S*s presented with the SK movie sequence also reported that the separated sexuality and affection to a greater extent than *S*s shown the KS sequence.

An insignificant correlation of .046 was obtained between the movie (absolute difference measure) and questionnaire measures of splitting.

DISCUSSION

The hypothesis that high authoritarians would overemphasize the sexual attributes of the star was confirmed. This result is consistent with the underlying theory of the authoritarian personality. Basic to the evolution of the personality structure of the authoritarian is the unacceptability of id impulses. Sexual strivings remain for the most part suppressed and isolated. However, since they are not integrated with the rest of personality, these impulses find projected and other displaced outlets, as in a heightened concern and moral indignation of sexual expression in others. One can then view the high authoritarian's essentially sexual perception of the star in terms of his projecting unacceptable sensual impulses on to her. However, another factor may be involved. Even when sexual impulses are acted upon by the authoritarian, it is rarely for the purpose, primarily, of genuine sexual gratification. For the male authoritarian, sexual expression tends not to be a goal in itself, but a means by which status is acquired and his sense of masculinity enhanced. His agreement with items suggestive of sexuality may, in part, be attributed to his desire to demonstrate a heightened sexual interest in women as proof to himself as well as others that he is fulfilling his idealized conception of the masculine role.

The hypothesis that high authoritarians would show a splitting of sexuality and affection was partially substantiated. Analysis of the heterosexual questionnaire scale supported the view that high authoritarians possessed dichotomized sex attitudes to a greater degree than low authoritarians. They should therefore find it more difficult, relative to low authoritarians, to react both sexually and affectionately toward the same woman. Instead, authoritarians show a tendency to react sexually toward women they do not respect or admire and inhibit sexual expression when conscious feelings of respect and affection are evoked. Findings on the organization of impressions, while not statistically significant, were in the predicted direction. The one inconsistency was that in their ratings (as determined by the absolute difference between each *S*'s sexual and kindness scores), high authoritarians were not found to separate sexuality and affection to a greater extent than low authoritarians. To obtain a high "splitting index" requires that *S* have a high score on one scale and a low score on the other. The rationale of this measure lies in the assumption that if sexuality and affection could not be integrated, *S* would have to acknowledge only one attribute, the other characteristic being incompatible and hence denied. However, the converse of this, whereby *S* acknowledges and responds to both characteristics, does not necessarily imply integration. Sexuality and affection can both be recognized, yet remain phenomenologically as isolated attributes. The absolute difference score hence does not indicate whether cognitive fusion actually occurred. What was found was that most *S*s did perceive both sexuality and affection in the star. While the high authoritarians did not show splitting, in the terms just outlined, to a greater degree than the lows, we cannot conclude that the same degree of integration was present for both groups. It seems that a more valid index of the integration of sexuality and affection is provided by the integrated category of the written impressions, which goes beyond the perceptually given and requires merging at a cognitive level.

Although analysis of the written impressions tended to support the hypothesis that high authoritarians would form more nonintegrated impressions, relative to the lows, this finding fell short of statistical significance. One possible explanation is that the star's behavior may not have been sufficiently sexual or kindly to represent extremes or stereotypes. It should be remembered that in the sexual scene, behavior suggestive of promiscuity or immorality was avoided. Again, in the kindness scene, it may have been possible for some to infer a sexualized quality in the star's behavior toward the adult male, even though affection was manifestly revealed. Further, it was impossible to portray only sexuality and kindness, which, while representing the major behavioral characteristics, were confounded to some extent with other attributes. Slovenliness or boredom was hinted at in the sexual scene and intelligence and cultivation in the kindness scene. In addition, the content of several of the items of the sexual scale implied a "looseness" beyond mere sexual attractiveness. *S*s thus may have been reacting to other qualities apart from the star's kindly and sexual features, even though inspection of the written impressions revealed that the great majority focused primarily on the two major attributes.

Viewed as a whole, the present findings substantiate one of the underlying hypotheses of *The Authoritarian Personality* in that high authoritarians were found to over-react to sexuality in women. While present results also support Frenkel-Brunswik's conclusions regarding the male authoritarian's tendency to separate sexuality and affection in actual behavior, the phenomenon of nonintegration was only equivocally demonstrated in the experimental task. The results are also significant in that they demonstrate the relevance of authoritarianism to men's social perception of women.

Summary

The present study investigated the effects of authoritarianism on men's reactions to sexuality and kindness in women. Sixty-four college males served as subjects, half the group consisted of high authoritarians and the other half comprised the low authoritarian group. The subjects were shown a silent movie consisting of two scenes, one sexual and one kindly, in which a young, attractive woman appeared. For half the group the sexual scene was shown first, and for the other half the kindly scene was first presented.

After seeing the movie, subjects wrote their impressions of the young woman's (star's) personality, after which they evaluated the star on a specially devised rating scale. The subjects then answered a questionnaire relating to their heterosexual behavior and attitudes.

The high authoritarian group perceived the star as significantly more sexual than kindly and also reported "splitting" of sexuality and affection in heterosexual behavior to a significantly greater degree than was reported by low authoritarians. The highs, in comparison to the lows, tended to form unfavorable impressions of the star, in which sexuality and kindness were not integrated. However, this difference was not statistically significant.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

Gollin, E. S. Forming impressions of personality. Unpublished doctoral dissertation, Clark University, 1951.

Gollin, E. S. Forming impressions of personality. *J. Pers.*, 1954, **23**, 65–67.

Gollin, E. S. Organizational characteristics of social judgements. A developmental study. *J. Pers.*, 1958, **26**, 139–153.

Gollin, E. S., & Rosenberg, S. Concept formation and impressions of personality. *J. abnorm. soc. Psychol.*, 1956, **52**, 39–42.

Jones, E. E. Authoritarianism as a determinant of first impression formation. *J. Pers.*, 1954, **23**, 107–127.

Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

Scodel, A., & Freedman, M. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92–95.

Scodel, A., & Mussen, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

Thibaut, J. W., & Riecken, H. W. Authoritarianism status and the communication of aggression. *Hum. Relat.*, 1955, **8**, 95–120.

(Received June 29, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 335-340

GROUP INTERACTION AND PERCEIVED SIMILARITY OF MEMBERS[1]

EZRA STOTLAND, NICKOLAS B. COTTRELL
University of Washington

AND GORDON LAING
University of Michigan

THE problem of the present study is the relationship between the amount of interaction in a group and the members' tendency to perceive other members as being similar to themselves. The general hypothesis is that members of a group in which there is a high degree of interaction perceive more similarity between themselves and other members than members of a group in which there is a low degree of interaction. The validity of this hypothesis is assumed to be limited to those cases in which the interaction conveys no information about the characteristics of the group members.

The theoretical source of this hypothesis is in the work of Michotte (1946). His studies of the perception of causal relationships between physical objects suggest that when one object is perceived to cause a change in a second object, the change in the second is perceived to belong to the first. In other words, a new unit or new entity is perceived, consisting of the first, causing object and of the change in the second object. This proposition has been extended theoretically to the area of social psychology by Heider (1958) who postulates that a person and an event he has been perceived to cause tend to be perceived as a single unit.

These ideas can be extended theoretically to the area of the perception of the interaction between two (or more) persons. By definition when two people interact, each act by one of them is followed by or causes some event (e.g., change in behavior) in the other. The causing person and the event in the other person should then constitute a perceptual unit. A series of such units should be formed as each person acts so as to cause an event (or reaction) in the other person. It is not unreasonable to assume that these series of units cumulate or synthesize to form a composite single unit, since each unit in the series is constituted by the same two (or more) persons. Such a synthesis would be expected from the laws of cognitive organization, which hold that, in general, the perceiver attempts to achieve the simplest possible organization of cognition. The composite units formed in interaction can vary in the number of interactions upon which they are based. Let us term this difference as the degree of strength of the composite unit, the larger the number of individual units, the greater the strength. The idea of strength is analogous to the concept of strong and weak gestalts.

One additional assumption is needed to derive the hypothesis of the present study. The phenomenon of assimilation found in the study of the perception of physical objects is the tendency to perceive all parts of the same perceptual unit as having the same quality, e.g., degree of brightness. It is reasonable to expect that this assimilative tendency toward uniformity of parts of a unit would be greater in strong units than in weaker units. As mentioned above, the number of interactions between two or more persons defines the strength of the resultant unit that they generate. Therefore, it would be expected that when there are many interactions between or among them, there should be a greater tendency to perceive them as uniform.

Perception of uniformity can arise in a given member of a group of two or more persons either by his perceiving that the other members have characteristics similar to ones he already believes he has, or by perceiving himself to have characteristics that he previously assumed the other persons to have. The present study is limited to the attribution of own characteristics to others. Our hypothesis is that in groups in which there are a high number of interactions among members, each member should attribute more of his own characteristics to the other members than in groups in which there are a low number of interactions. The underlying process is assumed to occur independently of the perception of similarity of members resulting from their acceptance of the group norm or standard, and independently of any perception of other members that may result from the interchange among the members of information regarding themselves.

---

[1] The work on this study was financed by a grant (M-2423) from the National Institutes of Health of the United States Public Health Service.

METHOD

Subjects. Ss were 90 students, 58 men and 32 women, required to participate as an assignment in an introductory psychology course at the University of Washington. In assigning *S*s to six-person experimental groups, an effort was made to assign *S*s from different living groups, discussion sections, majors, etc. to minimize the possibility of previous mutual acquaintance. Each group consisted of like-sexed persons.

Procedure. Ss arrived in the experimental room in groups of six. They were asked to sit around a small round table which had a Lazy Susan in the center with six compartments. Each *S* could see only the faces of the other *S*s, since there were wooden separations between them. In the compartment confronting each *S* was a pack of six slips of paper with 10 syllables on each one; a pack of six slips with five lines on each numbered from 1 to 5; a sheet with a title at the top and with two columns of 30 letters in all, in alphabetical order; a red cardboard; a pencil; and a pad of scratch paper.

*S*s were told that they were to list names from different geographical areas of the world, each person collecting names from a different area. Each *S* found the name of his area at the head of the sheet with the two columns of letters and also was given a list of all the areas (North America, Pacific, Latin America, Asian Mainland, European countries, European cities). There were world maps delineating these areas on the walls of the room. In addition, each of the compartments in the Lazy Susan was labeled for a given area.

*S*s were told that they were to form geographical names by combining syllables in pairs collected from sets of 10 syllables they found on the slips before them. Thus, each *S* produced five names by trying various combinations of the 10 syllables. He listed these names first on one of the sheets with five lines numbered 1 to 5 (although many *S*s did not obey the instructions to list the names first on the sheets). *S* was to write any geographical name from his own area on the sheet with the columns of letters, after the initial letter of the name he had formed. The objective was to find one name for each of the letters on his sheet. If *S* formed a name from an area other than his own, he was to write it on a slip of scratch paper, place it in the appropriate compartment in the Lazy Susan, and then turn the Susan until the right *S* received it. There was no trouble resulting from several *S*s desiring to use the Lazy Susan at the same time. Each *S* had six such sets of 10 syllables. Thus, any *S* could complete his collection of 30 names from names he had formed himself, as well as from names he received from others. If he found he was short a name after he had completed work on his syllables, he placed the red cardboard in his compartment so that other *S*s would notice that an error in sending a name had occurred, and correct it.

Few *S*s were actually able to complete their lists of 30 names, apparently because of the concoction of nonexistent names by some *S*s, lack of intellectual capacity, and failure to reject erroneous names received. Groups were allowed to continue until their functioning broke down as a result of the concoction of names. No group was allowed to run longer than 26 minutes on the task; the shortest time was 21 minutes. After data had been collected from 31 *S*s, a record was kept of the number of names completed by each remaining *S*, regardless of whether they were sent to other *S*s or returned. As will be shown below, the number of names completed had no effect on the dependent variable.

It is obvious that the task just described could be arranged so that each *S* would work mostly on names from his own area or mostly on ones from other *S*s' areas. In the *high interaction* condition, 25 of the 30 names each *S* formed from the syllables were from an area other than his own. He would have to send five names to each of the five other *S*s. He would, in turn, receive the 25 names he needed from the other *S*s. In the *low interaction* condition, 25 of the 30 names were from each *S*'s own area. He would send the other five to the other five *S*s, one to each, receiving one from each in return. The groups were assigned alternately to one or the other condition.

The task thus permitted variation in the degree of interaction among the *S*s, with about the same amount of work by each *S*. Each *S*'s contribution to the interaction could be equated, and the interaction conveyed no information about the *S*s' personalities.

When the task was completed, *S*s were given a list of 30 items from a variety of personality scales (e.g., general anxiety, hostility, personal history, test anxiety),[2] selected so as to refer to information about which students would not in general be likely to communicate with each other, and thus minimizing the possibility that *S*s would use their knowledge of typical student feelings and behavior in predicting other *S*'s responses. The *S*s first indicated whether each of these statements was true with regard to themselves. Then they were asked to respond to an identical list on a separate sheet, guessing how the majority of the other *S*s in the group would respond to each item. The dependent variable was the number of agreements between own responses and predictions of others' responses. Finally, *S*s filled out a post-experimental questionnaire with 7-point scalar items.

RESULTS

Before the hypothesis of the study can be tested, it is necessary to show that the high and low interaction conditions differed in no way (except in the degree of interaction) that could explain differences in the levels of the dependent variable. In Table 1 are presented the mean differences between the conditions in items of the post-experimental questionnaire. It can be seen there that the two conditions did not differ in degree of attraction to the group or to its members, interest in the various experimental tasks, perception of group "performance," or perception of integration into a "close-knit" group. However, there tended to be a higher degree of prior acquaintance among the *S*s in the low interaction condition than in the high interaction condition, a difference that must be taken into account in subsequent analyses.

---

[2] The authors wish to thank Irwin Sarason for his assistance in obtaining these items.

The hypothesis of the study was that in the high interaction condition, the *S*s would predict more similarity between themselves and the other *S*s than in the low interaction condition. There was no significant difference between the two conditions in this respect. For the high interaction condition the mean was 22.74, for the low, 22.53.

However, a clearer picture of the results is obtained by taking into consideration two other factors which can influence perception of own similarity to others, i.e., attraction to others and prior acquaintance with them.

TABLE 1

MEAN RESPONSES OF SUBJECTS TO ITEMS IN THE POST-EXPERIMENTAL QUESTIONNAIRE

Item	Experimental Condition	
	High Inter-action ($N = 45$)	Low Inter-action ($N = 44$)
1. Degree of prior acquaintance with other subjects[a]	7.27*	8.57*
2. Desire to get to know other subjects[a]	19.91	19.38
3. Interest in geography game	5.56	5.40
4. Interest in filling in personality questionnaire regarding self	4.29	4.18
5. Interest in predicting other subject's responses to personality questionnaire	4.23	4.00
6. Evaluation of group performance on geography game	3.75	3.56
7. Degree of integration of group into a close-knit group	3.44	3.18
8. Desire to leave present group to join another working on the same task	3.36	3.36

[a] Computed by having the *S* rate his degree of prior acquaintance with (Item 1) or his desire to know (Item 2) each of the other *S*s and summing up these ratings.

* Significantly different by a two-tailed *t* test, $t = 2.043$, $p < .05$. A high number indicates greater prior acquaintance. No other means were significantly different between conditions.

TABLE 2

MEAN NUMBER OF PREDICTIONS OF OTHER SUBJECTS' AGREEMENTS WITH SELF FOR SUBJECTS OF NO PRIOR ACQUAINTANCE

($N = 6$ in each cell)

Desire to Know	Experimental Condition		Total
	High Interaction	Low Interaction	
High	25.00	22.50	23.75
Low	22.17	19.57	20.87
Total	23.58	21.04	

Table 2 presents the mean number of predictions of similarity for *S*s in each experimental condition with no prior acquaintance with other *S*s, controlling on the desire to know the *S*s. The analysis of variance, shown in Table 3, indicates that experimental conditions, as well as the desire to know other *S*s, have highly significant effects on the number of predictions of similarity of other *S*s to self. In Table 2, it can be seen that the major hypothesis of the study is confirmed, since the mean for the high interaction condition is significantly greater than for the low. A similar comparison of the conditions for *S*s who had prior acquaintance produced no significant difference. The hypothesis holds only if the *S*s have no other information about the other *S*s because of prior acquaintance. Among these *S*s, the desire to know other *S*s is associated with prediction of similarity, as would be expected from other research.

However, it is possible for the findings to be explained by a process quite different from the one outlined above. The first step in the process would be for the *S*s in the high interaction condition to develop more liking for each other than those in the low interaction condition. The greater liking would then lead the *S*s in the high interaction condition to attribute more desirable attributes to the others. Since in general persons tend to attribute desirable traits to themselves, the outcome of this process would be greater similarity in the high interaction condition between the *S*s' self-descriptions and his descriptions of the others than in the low interaction condition. This is precisely the result obtained.

This explanation in terms of the influence of

TABLE 3

Analysis of Variance of Predicitions of Other Subjects' Agreements with Self for Subjects of No Prior Acquaintance

Source	*SS*	*df*	*MS*	*F*
Experimental Condition	42.666	1	42.66	11.91**
Desire to Know	54.000	1	54.000	15.07**
Experimental Condition × Desire to Know	.166	1	.166	
Within cells	71.668	20	3.583	
Total	168.500	23		

** $p < .01$.

interaction on attraction would, however, be difficult to reconcile with the results of Tables 2 and 3, since these tables show that there is more prediction of similarity in the high interaction condition than in the low, independent of the degree of attraction to other members. If the attraction explanation were valid, it would be expected that there would be no difference between the conditions independent of the degree of attraction.

Another way of resolving the issue between the "attraction" explanation and the one described earlier in the paper is to determine whether the *S*s in the high interaction condition tended more than those in the low interaction condition to predict that the other *S*s in their groups were similar to themselves with regard to socially desirable traits. If the difference between conditions occurred only with regard to the socially desirable traits, the attraction explanation would be supported. If the difference between conditions occurred equally for both desirable and undesirable traits, then support would be given to the explanation in terms of greater unit formation with greater interaction.

The following procedure was used to resolve this issue: 75 undergraduate *S*s of the same general background as those in the present study sorted the statements used in the personality questionnaire into six equal piles in the order of desirability of the statement's being true about other people. These rankings were pooled, and the total sample of statements was then split at the median rank on social desirability.

This split of the items on social desirability was then used to analyze the results of the original study. A socially desirable *response* was defined as an answer of "true" to a socially desirable statement or as an answer of "false" to a socially undesirable statement. For those statements to which an *S* gave a socially desirable response, the number of his predictions of similar responses by other *S*s was calculated. A similar total was calculated for those statements on which an *S* made an undesirable response about himself.

It is obvious that a person who attributes a high number of socially desirable traits to himself will have more opportunities to predict that others have the same socially desirable traits than will a person who attributes a low number to himself. Therefore, a count was made of the number of socially desirable responses each *S* made with regard to himself and the *S*s were split into two groups at the median. The mean numbers of predictions of similar socially desirable and undesirable responses were then calculated separately for the high and low social desirability *S*s. These means were calculated for the high and low interaction conditions, employing only those *S*s who had no prior acquaintance. The results are presented in Table 4.

This table shows no difference between the experimental conditions in the number of own socially desirable responses attributed to others, regardless of the number of own socially desirable responses. This lack of difference does not support the interpretation in terms of difference in attraction. On the other hand, a difference between the conditions significant at the .01 level is obtained for the prediction of socially undesirable responses made by *S*s low in the social desirability of their self-descriptions. In the high interaction condition, these *S*s had a greater tendency than in the low interaction condition to perceive the other *S*s as having the same undesirable traits as themselves. This difference is therefore completely inconsistent with the interpretation in terms of the effect of interaction on liking. Thus, both these data and the previously cited independence of the effects of attraction and of the experimental conditions argue strongly against the attraction interpretation. However, these results do not fit the "unity" explanation very well either.

The results just described appear to be inconsistent with Homans' (1950) hypothesis that greater interaction leads to greater liking. However, another test of the Homans hypothesis can be made by examining the difference between the conditions in attraction to the other members separately for *S*s high and low in the social desirability of their self-descriptions. In Table 5 are presented the appropriate mean post-experimental ratings of desire to know other *S*s, for the *S*s of no prior acquaintance. None of the differences between conditions is significant. Nonetheless, the general trend is consistent with the Homans hypothesis, especially for *S*s low in the social desirability of self-descriptions, who are more attracted to the other *S*s in the high interaction condition than in the low. However, these *S*s also predicted that the others were more like themselves in undesirable traits in the high interaction condition. It appears then that for these *S*s the similarity of traits was more significant for the degree of attraction to others than was the desirability of others' traits. Therefore, Homans' hypothesis may be valid with regard to attraction, but not with regard to favorability of perception.

TABLE 4

MEAN NUMBER OF PREDICTIONS OF SOCIALLY DESIRABLE AND UNDESIRABLE RESPONSES BY OTHER SUBJECTS SIMILAR TO RESPONSES WITH REGARD TO SELF FOR SUBJECTS OF NO PRIOR ACQUAINTANCE

(*N*s in parentheses)

	Experimental Condition	
	High Interaction	Low Interaction
I. *S*s high in attribution of socially desirable traits to self		
Desirable perception of self	16.8 (5)	16.2 (6)
Undesirable perception of self	6.00	6.50
Total	22.80	22.70
II. *S*s low in attribution of socially desirable traits to self		
Desirable perception of self	13.1 (8)	11.7 (6)
Undesirable perception of self	10.0 (A)	7.50 (B)
Total	23.1 (C)	19.2 (D)

Note.—Cells significantly different by two-tailed *t* test:
A vs. B ($t = 3.157$, $p < .01$).
C vs. D ($t = 3.900$, $p < .01$).

TABLE 5

MEAN RATINGS OF DESIRE TO KNOW OTHER SUBJECTS FOR SUBJECTS OF NO PRIOR ACQUAINTANCE

(*N*s in parentheses)

*S*s' Attribution of desirable traits to self	Experimental Condition	
	High Interaction	Low Interaction
High	21.1 (7)	18.4 (8)
Low	21.3 (6)	15.8 (9)
Total	20.1	17.5

Note.—No cells significantly different.

It was found that more names were composed by *S*s in the low interaction condition than in the high, the mean for the low interaction condition being 26.21 names, which was significantly different at the .001 level (t = 3.04) from the mean of 21.73 for the high interaction condition. (This mean is based only on the 59 *S*s for whom these records were kept.) The reason for this discrepancy probably is that the world-wide frame of reference necessary in the high interaction condition troubled some *S*s. The question may be raised whether the number of words composed by an *S* affected his prediction of other *S*s' agreement with him in the two experimental conditions. However, when the conditions were compared separately for *S*s above and below the median in number of names formed, these two groups of *S*s were found not to differ in their reaction to the condition differences.

DISCUSSION

The results showed that the experimental hypothesis, that the degree of perception of similarity of others to self is a function of the degree of interaction between self and others, was supported only among those *S*s who had no prior acquaintance with other *S*s. It seems that when the *S*s have some knowledge of the other *S*s, they use this information in their judgments of the other *S*s. The more subtle process of projecting own traits onto others, as a result of interacting with them, can occur only if such information is unavailable. In addition, the results show that the interaction leads to more perception of similarity mostly by *S*s who attribute relatively more socially undesirable traits to themselves, and then by perceiving the others to be similar to self on these undesirable traits. It may be that these *S*s are more reactive to differences in social situations as has been found in other studies (e.g., Stotland, Thorley, Thomas, Cohen, & Zander, 1957). However, it is difficult to understand why they perceived the other persons in the high interaction condition as being more similar to them in undesirable ways. These results raise questions about the adequacy of explaining the findings in terms of more perception of unity of the group arising from greater interaction, as was postulated in the original theory.

Other factors as well as interaction can, of course, increase the perception of similarity between self and other persons. In the present study, perception of such similarity was associated with a high attraction to other persons, a finding consistent with other research. Another factor which could lead to perception of similarity would be the development of a group norm. However, the development of a norm is contingent upon the exchange of information of some sort among the group members, which was of course impossible in the present experiment.

SUMMARY

It was hypothesized that persons with a high degree of interaction among themselves would perceive more similarity among themselves than persons with a low degree of interaction. It was predicted that this hypothesis would be supported when the interaction did not convey any information relevant to the degree of similarity among the subjects. This hypothesis was tested in a laboratory experiment. The findings were: (*a*) that the hypothesis was supported among those subjects with no acquaintance prior to the experiment; (*b*) that there was an association between perception of similarity to other subjects and desire to get to know them.

REFERENCES

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HOMANS, G. C. *The human group.* New York: Harcourt, Brace, 1950.

MICHOTTE, A. E. *La perception de la causalité.* Paris: Vocin, 1946.

STOTLAND, E., THORLEY, S., THOMAS, E. J., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55–63.

(Received July 2, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 341-347

VERBAL RATE, EYEBLINK, AND CONTENT IN STRUCTURED PSYCHIATRIC INTERVIEWS[1]

FREDERICK H. KANFER
Purdue University

RECENTLY, several attempts have been made to relate the momentary changes in subject's (*S*'s) anxiety level during the interview to structural speech characteristics (Dibner, 1956; Mahl, 1956), to GSR (Dibner, 1958; Dittes, 1957), and to respiratory activity (Goldman-Eisler, 1955, 1956). Panek and Martin (1959) also found a high degree of coincidence of speech disturbances and GSR during interviews. The primary purpose of these studies has been to determine the validity of a construct of momentary anxiety and its indices in order to understand interview behavior more completely. In the final analysis, demonstration of specific relationships of these indices to the momentary *content* of the interview would also lead to refinement of interviewing techniques and more adequate use of the interview as a method in personality research.

In a series of earlier studies, verbal rate was related to some experimentally controlled modification of *S*'s momentary emotional state, either by conditioning (Kanfer, 1958) or by selecting topics to which *S* was differentially sensitive (Kanfer, 1959). The earlier findings suggested that verbal rate increases with *S*'s present anxiety level and that a relatively stable relationship exists between verbal rate and content of speech. The present study attempted to extend the generality of the earlier findings by introducing a series of conceptually related external criteria which reflect *S*'s current and long-standing emotional state. For the present study a group of *S*s was selected, whose emotional adjustment was wider in range, less stable in its momentary content-related fluctuations, and more sensitive to separate content areas than the college population on which the earlier research was done. Since interview behavior is undoubtedly a function of the importance of the interview outcome in *S*'s current life processes, research with volunteer college students attenuates the effectiveness of any experimental variable relating to the complex emotional life history of *S*. Therefore, the present study used structured interviews which were imbedded in the routine admission procedures of a psychiatric hospital and were perceived by *S* as highly important in his present life situation.

The present study was designed to investigate several related hypotheses. The *first* concerned the relationship between verbal rate and content. In selecting the content for the interview it was predicted that any topic which is clearly of great emotional impact to all *S*s would result in significantly higher verbal rate than a topic which is not immediately related to *S*'s present emotional problems. It was hypothesized that a direct question about *S*s' reason for coming to the hospital would yield a significantly higher mean verbal rate than questions relating to other aspects of their life.

The *second* hypothesis dealt with the evaluation of *S*'s momentary anxiety by a concurrent nonverbal measure. If high verbal rate is related to increased momentary anxiety, then some nonverbal measure of *S*'s tension level during the discussion of various interview topics would show similar changes.

The human eyeblink has been postulated to reflect many changes in a person's overt and covert activity. Meyer (1953) proposed blink rate as a measure of muscular tension, and Doehring (1957) and Meyer, Bahrick, and Fitts (1953) have shown its relationship to some anxiety measures. Since it can also be recorded without affecting *S*'s interview behavior, this measure was selected for the present study.

The second hypothesis may now be stated more specifically. If the predicted changes in speech rate over topics are related to *S*'s momentary anxiety level, then blink rate is expected to show similar differences with the highest blink rate on the topic which arouses

---

[1] This study was supported in part by Research Grant M-2027 (C1) from the National Institute of Mental Health, United States Public Health Service. The writer wishes to acknowledge the contributions of Ann Durkee-Riggs who served as *E* and assisted in the execution of the study.

the greatest tension. Although high correlations between blink rate and verbal rate are not expected because these variables differ in their *specific* determinants and the measures differ in their sensitivities, the general pattern of rate changes within topics is expected to be similar.

It has been suggested earlier (Kanfer, 1959) that verbal rate may relate to *immediate* changes in *S*'s emotional state even though an anxiety measure which reflects a lasting personality attribute does not allow differential predictions of verbal rate for individual *S*s on unselected topics. Nevertheless, speech and eyeblink patterns may be associated with other personality attributes measured by psychological tests, psychiatric ratings, and diagnostic classification. In order to test this suggestion and to provide grounds for better conceptualizations in this area, the general relationship between verbal rate, eyeblink, personality test scores, and psychiatrist ratings was also explored.

METHOD

Subjects. Thirty-eight married female psychiatric patients between the ages of 21 and 36 served as *S*s in this study. The mean age was 28.7 years. All patients were assigned to the study immediately after admission to the hospital, except those who were too disturbed to take the test battery, whose admission diagnosis suggested organic factors, or who were started on treatment before they could be interviewed for this study.[2] The following diagnostic categories were represented among the *S*s: acute schizophrenia (12), paranoid schizophrenia (9), depressive reaction (3), dissociative reaction (2), anxiety reaction (3), obsessive-compulsive (2), unstable and inadequate personalities (6), and schizoid personality (1). Two *S*s were later eliminated from the study due to their inability to complete the interview.

Apparatus and Procedure. A test battery consisting of the Taylor Manifest Anxiety Scale (MAS), the Cornell Index, and individual MMPI forms was administered as part of the hospital testing routine to groups of three or four *S*s. On the same day or the following day *S*s were seen individually for the structured interview.

The interviews were conducted in a soundproof room, with *S* facing the experimenter (*E*) across a table and sitting next to a one-way vision screen. A microphone was on the desk. A modified desk clock faced *E*. In the observation room, an observer (*O*) monitored a tape recorder and a timing device which fed a 30-second marker into the recorder and simultaneously activated *O*'s chronoscope. This arrangement permitted synchronization of the recorded verbal material and the eyeblink count made by *O*. Onset of time periods was marked by *O* after *E* completed a question in the structured interview. An electric counter was used to record eyeblinks for 30-second intervals during *S*'s response to each topic.[3]

After *S* was seated in the interviewing room, *E* gave the following instructions:

> I am interested in understanding how you feel about certain things so that we will be better able to help you. I will need as much information and as accurate an account as you can give. Therefore, you can help me if you answer my questions by talking about everything which comes to mind when I raise a topic. Tell me about your attitudes and feelings and give whatever thought or example comes to mind. It is important that you omit nothing. I will not be able to answer any questions after you begin talking. OK? Do you have any questions before we begin?

Four warm-up questions were given, during which *S* was encouraged to talk continuously. The questions for the warm-up were selected on the basis of data from pilot interviews. On these questions *S*s talked freely and did not introduce discussion of their illness or other topics raised in the experimental interview itself. After it was clear that *S* understood the instructions, the following questions were asked and *S* was allowed to talk for at least 2 minutes. After each question *E* summarized *S*'s remarks and proceeded to the next question. For half of the *S*s, the order of Topics B and C was reversed. The questions were:

A. Where do you live? Tell me about it. How long have you lived there? What do you think of it? What is it like?

B. Tell me a little bit about your family, how you get along with them, and what they are like.

C. Do you feel that you generally have been pretty successful in handling members of the opposite sex, and that you are fairly attractive to men?

D. Do you think you can tell me a little about the trouble that brought you here?

The interviewing and eyeblink recording procedures were practiced with several *S*s in the pilot study. During the interview *E* refrained from any motor activity which could be interpreted as supporting, agreeing, or disagreeing. *E* made no verbal response of any kind during the 2-minute interval.

A rating scale with detailed instructions was given to each *S*'s psychiatrist before he had access to any test findings on the patient. The psychiatrists had no knowledge of the experimental design and were told only that each patient would be interviewed briefly. They were asked to indicate the degree of conflict which they saw in the patient in her relationship with the family, her relationship with members of the opposite sex, and her attitude toward her present illness. Specific examples were given for the extremes of a seven-step scale. In addition, the psychiatrist was asked to indicate on which of the topics the patient "would be most uncomfortable if she had to discuss it in an interview."

---

[2] The patients were interviewed at Larue Carter Memorial Hospital, Indianapolis. The author wishes to acknowledge the cooperation of Donald F. Moore and the hospital staff in contributing their time and facilities to the execution of the experiment.

[3] The author wishes to acknowledge the assistance of Helen Fowler who served as *O*.

TABLE 1
MEAN VERBAL RATES FOR TOPICS AND PERIODS
(In words 30 seconds)

Periods	Topics			
	A	B	C	D
1	55.81	55.47	48.64	59.33
2	53.78	57.92	48.52	63.35
3	47.18	56.08	43.23	58.22
4	42.95	51.94	40.81	58.61

The test batteries were then scored by hospital clerical personnel under the supervision of *E*.

RESULTS

Topics B and C were given as the second or third topic in the interview, alternated for half of the *S*s. To determine the effect of order, *t* tests were run on verbal rate and eyeblink comparing the two subgroups on each topic. The tests yielded no significant differences between the subgroups. Therefore, all *S*s were pooled in subsequent analyses.

Verbal Rate

Verbal output was measured in words/30 seconds on all topics and treated by analysis of variance. The mean verbal rates on all topics are presented in Table 1. The results of the statistical analysis are shown in Table 2. Verbal rate differed significantly from topic to topic for all *S*s. The highest verbal rate was obtained on Topic D (Illness) which had been selected a priori for its high anxiety-evoking potential for all *S*s. A Duncan analysis on the groups indicated that all means differed significantly from each other at the .05 level. The findings thus support the hypothesis that verbal rate varies with topics and that the highest rate occurs on a topic which is anxiety-arousing to all *S*s and which represents an immediate and realistic problem in the current interpersonal setting.

Verbal rate decreased significantly over periods *within* each topic (Table 2). The nonsignificant topic × period interaction suggests that this pattern of decline was similar for all topics. Overall verbal rate did not decrease simply as a consequence of increased time spent with the interviewer. In fact, Topic D, the last topic introduced, showed the highest verbal rate. In addition, as has been mentioned above, interchanging Topics B and C for half of the *S*s did not result in rate differences. Thus, the rate decrease seems to occur within each topic, and the rate may increase again at the start of a new topic.

Eyeblink Rate

O's reliability of counting blinks was checked in several pilot interviews against two other judges. The operational definition of a blink was based on a complete eyelid closure. The practice resulted in near-perfect agreement of judges' counts for segments of 2 minutes on the pilot interviews.

The summary of the analysis of variance on eyeblinks is presented in Table 3. The results show that blink did not differ significantly over topics for all *S*s, although the topic means array is in the expected direction. Mean blink rates per 30 seconds for topics were: Topic A, 8.95; Topic B, 8.90; Topic C, 8.72; and Topic D, 9.65. If the triple interaction is used as an error term, although use of this estimate results in a lowered F when the variance component for $T \times P \times S \neq$ zero, then the blink rate changes over topics among *S*s can be tested by $T \times S/T \times P \times S$. The significant F ($p < .01$) indicates that the topics differentially affected *individual* *S*s' blink rates. While some *S*s varied in blink rate over topics, all *S*s did not do so consistently.

Since Topic D had been specifically selected for its anxiety arousing potential, the second hypothesis was further tested by a *t* test for correlated measures between each *S*'s mean blink rate on the sum of Topics A, B, and C vs. Topic D. This analysis yielded a *t* value of 1.66, with 35 *df*, which, for a one-tailed test, falls between $p > .05$ and $p < .10$. The results do not give clear-cut support to the hypothesis

TABLE 2
ANALYSIS OF VARIANCE SUMMARY ON VERBAL RATE FOR 30-SECOND PERIODS

Source	*df*	*MS*	*F*
Subjects (*S*)	35	3956.15	
Topics (T)	3	5739.90	22.04**
Periods (P)	3	1781.30	5.77**
T × P	9	203.19	0.82
T × *S*	105	260.47	
P × *S*	105	308.76	
T × P × *S*	315	247.61	
Total	575		

** $p < .01$

that blink rate increases for all *S*s when the interview centered about their illness. However, the data suggest that *S*s show differential blink rates on topics, with the specific pattern determined by individual differences. Furthermore, Topic D tends to yield higher blink rates than the *average* rate on the other topics.

The mean blink rate decreased significantly over periods within each topic. Blink rate increased again whenever a new topic was introduced. The mean blink rates, summed over all topics, were: Period 1, 9.82; Period 2, 9.15; Period 3, 8.75; and Period 4, 8.65.

It was hypothesized that both verbal rate and blink rate are related to *S*'s tension level during the interview. Personality theory suggests that tension in interpersonal situations varies with diagnostic groups. Specifically, the more extensive the patient's disorganization and the greater his tendency to use withdrawal from the environment, the less is the expected tension in an interview. The present sample could be broken down by diagnosis into three groups: acute schizophrenics, paranoid schizophrenics, and others, composed mainly of neurotics and character disorders. The neurotic group would be expected to yield the highest total word rate and blink rate, and the acute schizophrenics to yield the lowest. This hypothesis was tested by an analysis of variance over the three diagnostic groups for total verbal rate and blink rate. The means are shown in Table 4.

For total verbal rate, a nonsignificant *F* ratio of 1.78 (2,33 *df*) was obtained, indicating no evidence that the diagnostic groups differed from each other in verbal rate. The eyeblink data were tested for homogeneity of variance in the groups by the $F_{\max}$ test (Walker & Lev, 1953). The group composed of neurotics and character disorders showed the greatest variability and the acute schizophrenics had the lowest variance. The resultant $F_{\max}$ of 2.5 was not significant, and an analysis of variance was run. The *F* ratio failed to show significant differences between groups.

TABLE 3

ANALYSIS OF VARIANCE SUMMARY ON EYEBLINK RATE FOR 30-SECOND PERIODS

Source	*df*	*MS*	*F*
Subjects (*S*)	35	419.68	
Topics (T)	3	24.29	0.88
Periods (P)	3	47.39	3.77*
T × P	9	12.73	1.19
T × *S*	105	27.70	2.58**
P × *S*	105	12.58	1.17
T × P × *S*	315	10.74	
Total	575		

* $p < .05$.

** $p < .01$.

TABLE 4

TOTAL VERBAL AND BLINK RATE MEANS AND STANDARD DEVIATIONS BY DIAGNOSTIC CATEGORIES

	Acute Schizophrenics ($N = 11$)		Paranoid Schizophrenics ($N = 9$)		Neurotics ($N = 16$)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Verbal	820.55	219.96	841.56	233.60	853.81	272.16
Eyeblink	119.64	59.19	152.33	70.25	158.13	94.12

Relationship between Psychiatric Ratings, Test Scores, Verbal Rate, and Eyeblink

Correlations were computed between raw scores on eight of the MMPI scales, the MAS, the Cornell Index, and the psychiatric ratings, and separate verbal rate and eyeblink measures for periods, topics, and verbal and blink rate totals over the interview. Only those correlations which relate the dependent variables, i.e., verbal rate and blink rate, to the various criteria are discussed here. The values of the correlations are presented in Table 5.

Examination of the correlation matrix revealed that all MMPI measures, except the *Hy* and *Ma* scale, showed some relationship to verbal rate during the initial 30 seconds of talk on each topic. The direction of these relationships suggests that *S*s who give a large number of deviant responses on the MMPI scales show a lower initial verbal rate in the interview.

A second cluster of correlations can be seen in Table 5 between psychiatric ratings and blink rate. *S*s who were judged to show more severe conflicts tended to show a higher blink rate. Ratings on family conflicts and anxiety over present illness showed this tendency more consistently over topics and periods than the ratings of sex conflicts. The correlations do not appear to be greater when the ratings in a given area are compared with the behavior on

TABLE 5

CORRELATIONS BETWEEN TEST SCORES, PSYCHIATRIC RATINGS, AND INTERVIEW BEHAVIOR

Criterion	Eyeblink Rate									Verbal Rate								
	Topics				Periods					Topics				Periods				
	A	B	C	D	1	2	3	4	Total	A	B	C	D	1	2	3	4	Total
Hy										−216				−229				
Hs										−244	−219			−359**		−221		
D		231	305*	208			253			−333**		−251		−408***		−200		
Pa										−223	−287*	−203	−206	−426***		−237		
Pt														−364**				
Sc														−328*				
Ma																		
Si			209				205			−310*				−381**	−245			−226
Rating (Family)	276	290*	292*	282*	329*	249	342**	296*	311*									
Rating (Sex)	234	332**	306*			276	300*	238	272					250			214	
Rating (Illness)	284*		309*	303*	356**	312*	242	219	284*	271						225		
MAS														−260				
Cornell																		

Note.—All correlations between −.20 and +.20 are omitted from this table. Decimal points omitted.

* $p < .10$.

** $p < .05$.

*** $p < .02$.

the topic in the same area, e.g., when ratings on family conflicts are compared with Topic B.

The psychiatric predictions of topics on which *S*s would be most and least uncomfortable were compared with the verbal rate and blink rate on these topics. Arcsin transformations were used to obtain *z* values for an estimate of the probabilities associated with these predictions (Walker & Lev, 1953). The *p* values represent the significance levels for the hypothesis of chance success ($p = .33$) in predicting Topics B, C, or D. Prediction of greatest discomfort on a topic coincided with *S*'s highest verbal rate on the same topic in 14 *S*s ($p = .231$). Lowest verbal rate coincided with prediction of least discomfort in 9 *S*s ($p = .14$ below expected chance value). *S*'s highest blink rate coincided with prediction of greatest discomfort in 16 *S*s ($p = .09$). Lowest blink rate coincided with prediction of least discomfort for that topic in 15 *S*s ($p = .13$).

DISCUSSION

The results confirm the first general hypothesis, which stated that verbal rate is related to content. They also support the prediction that verbal rate increases as the emotional impact of a topic increases. The present results gain in significance when related to earlier findings (Kanfer, 1959) from similar interviews with a college population With college students, verbal rates among topics differed only when within-subject changes were considered and adjustment ratings were used to determine the topic on which the greatest momentary anxiety was expected to occur in the interview. In these *S*s, the topics did not seem to evoke similar emotional responses, and therefore rates varied over topics. In the present psychiatric population, the topic relating to hospitalization appeared to be uniformly anxiety-arousing in all *S*s. As the period means show (Table 1), verbal rate in Topic D started higher and stayed higher when compared with the other topics.

The lowest verbal rate was obtained on Topic C, dealing with *S*'s heterosexual attitudes. This finding is somewhat surprising, in view of the etiological importance which this topic might be assigned in the present population and the earlier findings of the high correlation between verbal rate and low adjustment on this topic with college students. A major reason for this result may be the different way in which the two populations handled this question. The late adolescent, unmarried students talked about dating, their sexual attitudes, and their conflicts and fears in this area. The older, married patients talked mostly about their husbands and generally expressed positive feelings toward them. None of the patients discussed any sexual material. It is possible that the patients were evading any deep sexual problems. Presumably this would enable them to reduce their momentary anxiety in the interview.

The *family* topic (B), shows the second highest rate. This result is compatible with the findings of the highest correlation between verbal rate and adjustment ratings ($r = .467$) in the college student data.

The eyeblink data do not yield support for the second hypothesis. However, the data suggest that some common factor, such as the postulated momentary anxiety level, may partially underlie both behavioral changes even though these measures may not be highly correlated. In comparison with verbal rate, blink rate may be more sensitive, so that the special pertinence of unique events recalled during discussion of a topic may result in less uniformity over topics for all *S*s. The finding of a lower average blink rate on Topics A, B, and C certainly suggests that there was a general trend toward higher anxiety arousal on Topic D. The within-topic changes in blink rate parallel the pattern found on verbal rate. If momentary tension decreases as discussion of one topic continues, then both measures reflect this change in the predicted direction.

In the correlation matrix in Table 5, the size of the *r*'s must be evaluated in relation to the small *N*, which requires a high degree of relationship for significance. The distribution of *r*'s (Table 5) shows two distinct clusters of correlations. The self-inventory scores relate consistently to verbal rate during the initial 30-second periods. The psychiatric ratings, derived from behavioral observations and life histories, show a consistent, although low relationship to blink rate. In view of the moderate correlations between the variables reported here, due caution should be exercised by viewing these data as a basis for further research hypotheses; the correlations cannot be interpreted as validating relationships which have practical utility in clinical diagnosis.

The results suggest several post hoc speculations. The initially low verbal rate in MMPI high scorers may reflect a general inability to focus and organize their efforts at the start of a new "problem," i.e., a new topic. However, once a start is made, speech may come more easily under control of momentary tensions; hence, the correlations disappear. This is indicated by the decrease of correlations over periods within topics, and over topics as the interviews progressed.

The direction of the correlations between MMPI scores and verbal rate suggests that responsiveness in the interpersonal setting is greater in *S*s who show fewer specific symptoms or whose illness has not yet crystallized. Those *S*s who were most disturbed in their interpersonal relationships also tended to show the lowest verbal rate in the interview, possibly because they were least affected by *E*'s presence and unaware of any inappropriateness in their behavior. In this sense, their symptoms may well serve as efficient defenses in the interview. Although the dynamics of the depressive and the paranoid vary, the high correlations with verbal rate appear meaningful in both cases. The motor retardation in the depressed and his hypothesized repressed hostility towards others may reduce verbal rate as the severity of the illness increases. The paranoid patient's suspiciousness may similarly decrease his verbal output as symptoms increase.

Considering the frequent incongruence between psychiatric ratings and behavioral criteria, the cluster of correlations between ratings and eyeblink presents an interesting finding. It may be speculated that ratings by psychiatrists are often based on observation of somatic behaviors. Thus, the ratings would not be independent of *S*'s blink rate. However, such reasoning supports the pertinence of blink rate as a somatic concomitant of conflict behavior.

The overall results of this study provide further support for the existence of a relationship between verbal rate and verbal content in structured interviews. The data also suggest that during ongoing speech any verbal rate changes may relate to momentary tension level, to the extent that these changes are consistent with other components of the total patterning of *S*'s emotional state. The present study also demonstrates the utility of the experimentally structured interview in the evaluation of various dimensions of patient behavior. The method allows considerable control over *E*'s behavior without sacrificing the richness or "natural" setting of the interview. In view of the rapidly growing evidence on the systematic effects of the interviewer's role on the patient's productions some controls over E's behavior seem essential in interview research.

SUMMARY

The purpose of the present experiment was to investigate the relationship between verbal rate, content, and eyeblink rate in experimental interviews, and to provide further evidence for an earlier hypothesis that momentary anxiety increases verbal rate.

Thirty-six female psychiatric inpatients were given personality tests and were then seen for an experimental interview. The interview consisted of four questions to which the subject replied without interruptions for 2 minutes. The sessions were tape recorded and eyeblink rate was simultaneously recorded by an observer through a one-way mirror. The questions dealt with the following topics: (A) the present home, (B) family, (C) sexual attitudes, (D) reasons for hospitalization. Verbal rate was expected to be highest on Topic D, since this topic was selected a priori to be the most anxiety-arousing to the patient. In addition, each attending psychiatrist rated the patient with regard to degree of conflict in these areas.

Verbal rate and eyeblink rate were measured for 30-second intervals. MMPI scores, Taylor anxiety scores, and Cornell scores were also available for each *S*. Verbal rate was found to differ significantly over topics, with the highest rate on Topic D. Furthermore, there was a significant decline in speech rate within each topic, measured by comparing 30-second periods over all topics. The eyeblink data showed a similar decline within topics. While mean blink rate did not differ among topics, a significant topic × subjects interaction indicated that the subjects' blink rates were affected differently by the topics. A *t* test for correlated measures between average blink rate on Topics A, B, and C vs. D suggested that *S*s tended to yield higher blink rates on Topic D than on the average of the other topics.

Correlations between personality test scores, psychiatric ratings, and eyeblink and verbal rate revealed two clusters of moderate correlations. The first cluster contained significant correlations between high scores on five of eight MMPI scales and low verbal rate during the initial 30 seconds on each topic. The second cluster contained significant positive correlations between psychiatric ratings of degree of conflict and eyeblink measures.

The results supported the hypothesized relationship between verbal content and verbal rate. The eyeblink findings were considered as further evidence of the relationships between momentary tension, verbal rate and verbal content. The relationships of the dependent variables to the test scores and psychiatric ratings were discussed as suggestive of the importance of personality dynamics as determinants in the complex process of interpersonal communications.

REFERENCES

DIBNER, A. S. Cue counting: A measure of anxiety in interviews. *J. consult. Psychol.*, 1956, **20**, 475–478.

DIBNER, A. S. Ambiguity and anxiety. *J. abnorm. soc. Psychol.*, 1958, **56**, 165–174.

DITTES, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, **54**, 187–191.

DOEHRING, D. G. The relation between manifest anxiety and rate of eye blink in a stress situation. *USN Sch. Aviat. Med. res. Rep.*, 1957, No. 6. (Project NM 130199)

GOLDMAN-EISLER, FRIEDA. Speech-breathing activity: A measure of tension and affect during interviews. *Brit. J. Psychol.*, 1955, **46**, 53–63.

GOLDMAN-EISLER, FRIEDA. Speech-breathing activity and content in psychiatric interviews. *Brit. J. med. Psychol.*, 1956, **29**, 35–48.

KANFER, F. H. Effect of a warning signal preceding a noxious stimulus on verbal rate and heart rate. *J. exp. Psychol.*, 1958, **55**, 73–80.

KANFER, F. H. Verbal rate, content, and adjustment ratings in experimentally structured interviews. *J. abnorm. soc. Psychol.*, 1959, **58**, 305–311.

MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1–5.

MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.

MEYER, D. R., BAHRICK, H. P., & FITTS, P. M. Incentive, anxiety and the human blink rate. *J. exp. Psychol.*, 1953, **45**, 183–187.

PANEK, D. M., & MARTIN, B. The relationship between GSR and speech disturbances in psychotherapy. *J. abnorm. soc. Psychol.*, 1959, **58**, 402–405.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

(Received August 10, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 348–354

THE SUSCEPTIBILITY OF CHRONIC SCHIZOPHRENICS TO SOCIAL INFLUENCE IN THE FORMATION OF PERCEPTUAL JUDGMENTS[1]

CARMI SCHOOLER[2]
National Institute of Mental Health

AND

HERBERT E. SPOHN
Veterans Administration Hospital, Montrose

LACK of participation in social activity is such a prominent feature of the behavior of the chronic schizophrenic that it has led such diverse theorists as Cameron (1947), Fenichel (1945), and Fromm-Reichman (1939) to the general conclusions that the schizophrenic process involves a decline of responsiveness to social stimuli and that the degree of this impairment in schizophrenics varies with severity of illness.

The present study represents an attempt to test this general hypothesis experimentally. Hovland, Janis, and Kelley (1953) found that "persons who are most highly motivated to maintain their membership (in a group) tend to be most susceptible to influence by members of the group" (p. 276). If susceptibility to the influence of others is taken as an index of an interest in the social environment and responsiveness to social stimuli, it may be postulated that the degree of susceptibility to influence by others is inversely related to the degree of illness in chronic schizophrenics. The group situation first developed by Asch (1956) to study conformity behavior lent itself well to the experimental test of this hypothesis. In this procedure the subject (*S*) judges which of three comparison lines is the same length as a standard line after a group of the experimenter's confederates have given a wrong answer. The naive *S* is thus confronted by a choice between maintaining independence of judgment or conforming to the "norm" of the group. It was predicted that in the Asch situation regressed *S*s would conform less than partially remitted ones, who in turn would conform less than normal ones (Hypothesis I).

Two other variables were also investigated. One variable was introduced in order to study the possible effects on the hospitalized schizophrenic of the lack of contact with normals that characterizes his usual hospital environment. A previous study of the authors (Schooler & Spohn, in press) suggested that hospitalized schizophrenics accept degree of health as a prime criterion in selecting their social contacts and that their experience with the unreliability and inappropriateness of their peers' behavior also teaches them to distrust their peers' judgments. Consequently, it was predicted that schizophrenic *S*s to whom the confederates are introduced as normals would conform more than those to whom the confederates are introduced as fellow patients (Hypothesis II).

The third variable studied derives from the distinction made by Deutsch and Gerard (1955) between "normative" and "informative" sources of influence in the conformity situation. A person may conform in the Asch situation because he wishes to satisfy the positive expectation of the others (normative) or because he regards the others' judgments as superior to his own (informative). While in the standard Asch situation these two sources of motivation are implicit, they can, of course, be induced explicitly by instruction and thus function as two distinct additional sources of influence intensifying pressure to conform. No sound basis existed for predicting whether chronic schizophrenics respond selectively to these two sources of motivation to conform. It was of interest, however, to determine whether increments in pressure to conform would break through the hypothesized resistance to social influence in chronic schizophrenics. Specifically, it was predicted that chronic schizophrenics tend to conform more frequently when normative or informative motivation to conform is explicitly induced than when this is not the case (Hypothesis III).

---

[1] This report is based on a doctoral dissertation submitted by the senior author to the Graduate School of Arts and Sciences, New York University, February 1959. The senior author wishes to express his appreciation to Murray Horwitz for his guidance in the course of the study.

[2] This study was carried out while the senior author was a social psychology trainee at the Montrose Veterans Administration Hospital. The authors wish to express their gratitude to the staff of that hospital and to the staff of the Castle Point Veterans Administration TB Hospital. Special thanks are due to Jacob Cohen for his invaluable aid in dealing with the statistical problems which arose. A portion of this study was presented at the APA meeting in Washington, D. C. in August 1958.

A series of hypotheses were also projected about the interactions among the several variables. These hypotheses were based on the reasoning that if degree of illness is inversely related to responsivity to social stimuli it would also be inversely related to readiness to respond differentially to relevant differences among social stimuli. Therefore, it was predicted that healthier *S*s tend to be more responsive to differences in the nature of the social pressure (Hypothesis IV-A) and the apparent mental health status of the confederates (Hypothesis IV-B).

METHOD

Subjects

One hundred and forty-four *S*'s were tested: 48 regressed schizophrenics, 48 partially remitted schizophrenics, and 48 hospitalized TB patients who served as normal controls. The regressed *S*s were chosen from the hospital's closed wards by clinical psychologists who used the Montrose Rating Scale (Rackow, Napoli, Klebanoff, & Schillinger, 1952–53); partially remitted *S*s were selected from the open wards through the same procedure. Only those *S*s whose score was below 10 on this scale were considered regressed and only those who were scored above 17 were considered partially remitted, these scores representing points on the scale that had been established previously as reliable cut-off points (Rackow et al., 1952–53). All of the schizophrenic *S*s had a history of at least 2 years of continuous hospitalization prior to the time of testing. The normal *S*s were chosen from among the patients of a nearby Veterans Administration TB hospital. None of those used had any known psychiatric history. All of the cells of the experimental design were equated for age and education (see Schooler, 1959). The mean age of the *S*s was 33.9 years; their mean level of education 10.2 years.

Design

The schizophrenic *S*s were tested under a 2 × 2 × 3 factorial design that permitted the following comparisons: partially remitted *S*s versus regressed *S*s, confederates introduced as normals versus confederates introduced as patients, standard Asch situation versus informative influence versus normative influence. Each of the 12 cells contained 8 replications. The normal controls were tested under the informative, normative, and standard Asch situations with 16 replications in each cell. For these *S*s the variable pertaining to mentally ill peer-confederate versus normal-confederate was not introduced on the grounds that the meaning of this variable would not be comparable. For TB patients the distinction between peers and nonpeers could only have been made on the basis of institutional status, while the major concern in manipulating the variable among the original population of schizophrenics was to isolate the effects of differences in the mental health status of the confederates.

Procedure

The series of test cards, their order of presentation, and the confederates' responses duplicated Asch's procedure (1956). All *S*s were also pretested in an individual session several days before the experiment in order to insure that they were able to make all of the required comparisons correctly. Furthermore, schizophrenic *S*s were retested in an individual posttest immediately after the experimental situation.

In the actual experimental situation each *S* was tested with three confederates. The apparent mental health status of the confederates was varied by dressing the confederates who were to be presented to the schizophrenic *S*s as mentally ill patients in typical patient garb and introducing them as coming from wards housing patients of the same health level as the *S*. When the confederates were presented as normals, they wore the uniforms of their actual status—that of dietary worker.

The informative condition was implemented by asking each member of the group immediately before the test began, how well he did when he took the test earlier. Each of the confederates replied with a prearranged sentence which stated that he had made all of the comparisons correctly.

The normative condition was presented by having the confederates express approval of unanimity and disapproval of disagreement through a series of preprogramed comments, which were made the first three times the *S* agreed with the majority and the first three times he disagreed.

In the course of testing the normals it became apparent that many of the *S*s knew the confederates personally. Since this was not the case with the schizophrenic *S*s, it was necessary to gather information about the degree of acquaintance between the normal *S*s and the confederates in order to evaluate the effects of this difference. Consequently, the normal *S*s were asked to rate each of the others with whom they had taken the test on a 5-point scale:

> 1. never saw him before, 2. saw him around but do not know his name or know him enough to speak to, 3. know his name and know him well enough to say hello to, 4. have occasional conversations with him, and 5. he is a fairly good friend of mine.

Method of Analysis

The original plan called for the use of analysis of variance. Since, however, the number of conforming errors was quite low (57% of the *S*s made no errors), the means of various cells were highly correlated with their variances (.98) and rank-order nonparametric statistics had to be employed. The Mann-Whitney *U* test was used for all two-condition analyses and the Kruskal-Wallis test for all three-condition analyses. Possible interactions between variables were explored by a method developed by Jacob Cohen. In this method the error scores of *S*s undergoing one condition of a given variable were randomly paired with the error scores of *S*s undergoing the second condition of that variable, the other variables in the experiment being held constant. The differences between these paired scores were then ranked. These ranks were then summed separately for each of the conditions of the second variable and the sums compared through the Mann-Whitney or Kruskal-Wallis tests. Although this procedure does not represent a unique solution, since only one of many possible sets of random pairings was examined, the method was considered an adequate device for exploring interactions, especially since the fact that a large number of *S*s produced no errors reduced the likelihood that a different random pairing would have produced different results.[3]

Several other problems arose, especially in the analyses of the data of the normal *S*s. First of all, a correlation of .51 was found between the normal *S*s' rating of acquaintance with the confederates and the amount of conformity shown. Consequently, it was decided that only those normal *S*s would be included in the analyses whose degree of acquaintance with the confederates was similar to that existing between the schizophrenic *S*s and the confederates with whom they were tested. Only those normal *S*s were included who indicated by their scale responses that they did not know the names or have a speaking acquaintance with the confederates. Eight patients who met these criteria were randomly selected from each condition.[4]

Another problem that arose in analysing the responses of the normal *S*s concerned the selection of that group of schizophrenic *S*s who could appropriately be compared with the normals. Normal *S*s, as noted above, could not be administered the treatment variation of peer-confederates versus normal-confederates since the association between mental health classification and status level differs for normals and schizophrenics. Since conclusive arguments could not be made for either alternative, comparisons with schizophrenics in both the peer-confederate and normal-confederate conditions were made. In addition, the responses of the schizophrenic *S*s were analysed independently. Where no differences were found between the peer-confederate and normal-confederate conditions, the pooled data of all the schizophrenic *S*s were also compared with those of the normals.

RESULTS

Original Hypotheses

Using the nonparametric methods described above, the three hypotheses concerning main effects[5] stemming from degree of mental health (I), status of confederates (II), and source of motivation (III) upon conforming behavior in chronic schizophrenics were tested. None of the differences in frequency of conforming responses between groups or conditions was found to be significant. Thus, the three major hypotheses of this study were not confirmed.

The hypotheses predicting second-order interactions resulting from differences between regressed and partially remitted schizophrenic *S*s in selectivity of response to the several social discriminanda represented by the experimental conditions were similarly tested and not found significant. An exploratory analysis of the triple interaction between health, status, and nature of social pressure yielded an H value of 5.890 (using the Kruskal-Wallis test) significant at the .05 level.[6]

As a result of the analysis of the triple interaction two tentative conclusions were reached. First of all, the variables studied seem to affect the number of conforming errors made only if they are introduced in a particular combination. Secondly, since the regressed *S*s made proportionately more errors than did the partially remitted *S*s in the normal-confederate informative condition as opposed to the peer-confederate informative condition, it seems that the effect of the informative condition is more dependent upon the mental health status of the majority among the regressed *S*s than among the partially remitted ones. This difference suggests that compared with partially remitted patients, regressed patients give less weight to their peers as expert sources of information than they do to normals.

Nonpredicted Responses

Asch (1956) assumed that all the errors made by his *S*s reflected conformity to the responses of the majority. This assumption could not be made for all of the errors made by the schizophrenic *S*s.

Besides the conforming errors, several types of responses occurred during the critical trials which did not seem to directly reflect conformity to the responses of the majority. One such type occurred in those trials in which the third alternative line on the card, choice of which represented neither the correct answer nor the wrong answer given by the confederates, was not intermediate in length between the majority response and the correct answer. For example, in Trial 13, the correct choice was of a line 5 inches long, the majority chose a line 4 inches long, and the third alternative line was 6½ inches long. The choice of the third alternative in such a situation could not be construed to represent either an accurate response or a tendency to be influenced in the direction of the group. Such responses were labeled "noncorresponding" responses.

---

[3] For a more complete description of the method and a description of the similar method through which triple interactions were analyzed see Schooler (1959).

[4] These new cells remained matched for age and education with the other cells (see Schooler, 1959).

[5] The data in both raw and ranked form are presented in Schooler (1959).

[6] The value of H is significant at the .05 level of confidence if the beta approximation of the probability of a given value of H occurring is used. It does not reach the .05 probability value if either chi square or alpha approximations are used.

A second such type of error occurred in those trials in which the length of the third alternative line lay between the correct response and the one given by the confederates. Although Asch (1956) considered the choice of the intermediate alternative in such a situation a form of "partial conformity" to the norms of the group, the occurrence of the noncorresponding responses of the type just described casts some doubt on this assumption. It is conceivable that the choice of the intermediate line is based on the same motivation that led to the choice of the noncorresponding responses.

Two other types of errors also occurred besides those that occurred in the critical trials. One type, which was labeled "correct-wrong errors," occurred when the *S*s chose incorrect lines in those trials in which the confederates responded correctly. The other type occurred when *S*s made mistakes during the individually administered posttest.

This profusion of various types of errors raised the problem of combining the types into broader classes. One possible way of doing this was suggested by the fact that two distinct classes of errors could be distinguished on a priori grounds: the first class, nonconforming errors, including noncorresponding and correct-wrong errors, being made up of those errors which conform to neither the social pressure of the group nor the physical properties of the stimulus; the second class consisting of actual conforming errors.

In order to determine whether these two classes, distinguishable on a priori grounds, do in fact constitute two distinct empirical classes, the data of the regressed *S*s, the group that made the overwhelming number of nonconforming errors, were examined and intercorrelations among the various types of errors were obtained. The correlations shown in Table 1 support the assumption of two classes of errors. The correct-wrong and noncorresponding errors proved to be highly correlated with each other and not significantly correlated with the conforming errors. This finding seemed to justify a separate analysis for each of the classes.

TABLE 1

INTERCORRELATIONS AMONG TYPES OF ERRORS FOR THE REGRESSED SUBJECTS

Type of Error	Correct-Wrong	Noncorresponding	Intermediate	Posttest	Conforming
Correct–Wrong	—	.85*	.70*	.63*	.27
Noncorresponding		—	.86*	.71*	.20
Intermediate			—	.68*	.21
Posttest				—	.41*
Conforming					—

* Significant at .01 level.

The table of intercorrelations also makes it readily apparent that the intermediate errors, at least among regressed schizophrenics, are empirically a part of the class of nonconforming errors. Thus, it would appear that at least for regressed schizophrenics, such errors do not represent partial conformity. However, since this type of error could not be categorized unambiguously on a priori grounds and might have a different meaning when made by *S*s in other health categories, its analysis was not combined with that of the nonconforming errors. A separate analysis was also performed for the posttest errors.

The analysis of the occurrence of nonconforming errors among the schizophrenic *S*s revealed that the occurrence of such errors is directly related to degree of illness ($z = 3.156$, $p < .01$) and the presence of conditions designed to increase the amount of pressure ($z = 2.313$, $p < .05$). Since the great majority of nonconforming errors was made by regressed *S*s, it would be fair to say that when such *S*s are subjected to increased pressure towards conformity they respond by making a greater number of nonconforming errors.

Since in the analysis of the data from the schizophrenic *S*s neither the peer-confederate, normal-confederate variable, nor any of the interactions of which it was a part approached significance, it was possible to evaluate the data from the normal *S*s with the combined data from the schizophrenics from both the peer-confederate and normal-confederate conditions. This analysis showed that both degree of health ($H = 11.313$, $p < .01$) and the presence of conditions designed to increase the amount of social pressure applied ($z = 2.43$, $p < .02$) affected the number of nonconforming errors made. The difference between the health conditions was primarily due to the far greater number of nonconforming errors made by the regressed *S*s as opposed to the normal and partially remitted ones.

The analysis of the distribution of intermediate errors among the schizophrenic *S*s indicates that the occurrence of such errors is significantly affected only by degree of health ($z = 2.779$, $p < .01$). Since the peer-confederate–normal-confederate variable produced no significant main effects or interactions among the schizophrenics, it was possible to analyze the normal *S*s' data with the combined data of the schizophrenic *S*s from both the peer-confederate and normal-confederate conditions. This analysis showed that the production of intermediate errors was definitely related to degree of health, the regressed *S*s making many more such errors than the partially remitted or normal ones ($H = 11.030$, $p <.01$). However, no other significant differences were found.

The analysis of the posttest errors revealed that the occurrence of these errors was significantly related to degree of health, with the regressed *S*s making many more such errors than the partially remitted ones ($z = 2.645$, $p < .01$). A significant interaction was also found which indicated that more posttest errors were made after the normal-confederate informative condition than the peer-confederate informative condition ($H = 8.139$, $p <$ .02).

DISCUSSION

The results of this study do not confirm its basic hypothesis. Insofar as conforming behavior is an index of degree of responsiveness to the social environment, chronic schizophrenics are no less responsive to the influence of others than normals, nor are the severely ill less responsive than chronic schizophrenics who have attained some degree of recovery. The findings suggest that the manifest gross social withdrawal which characterizes many chronic schizophrenics *behaviorally* does not appear to be associated with a commensurate loss of sensitivity or awareness of others as social objects. Indeed, the results suggest that the hypotheses of this study should have been framed in reverse.

The significantly greater number of nonconforming errors made by the regressed *S*s suggests that as a group such *S*s are actually more reactive to the social stimuli of the Asch situation, albeit in an irregular way, than partially remitted and normal *S*s. Furthermore, both the effect of a particular combination of conditions in increasing the number of conforming errors made by the regressed *S*s and the increase of nonconforming errors under conditions designed to increase the amount of pressure suggest that such *S*s react differentially to the various interpersonal implications of the situations confronting them.

This possibility is of considerable theoretical interest since, by and large, as the hypotheses of this study suggest, most interpreters of chronic schizophrenia have postulated a diminution of social drives to account for the socially withdrawn behavior in chronic schizophrenics.

The nature of the motivation that underlies the nonconforming errors cannot be discerned from the data at hand. Two general hypotheses, consistent with the data, suggest themselves. On the basis of the high positive correlations between those types of errors (noncorresponding and correct-wrong) making up the class of nonconforming errors and the posttest errors, the type which seems most clearly to reflect disruption, one may postulate that the nonconforming errors occurred as part of a random pattern of responses reflecting disruption in the *S*s' cognitive functioning due to the stress of the experimental situation.

Such an explanation would be congruent with the conclusions of Rodnick and Garmezy (1957). The findings they report indicate that

> although schizophrenic patients can and do respond adaptively in tasks of considerable difficulty and complexity . . . this adaptability is a tenuous one which can be disturbed by the introduction of minimal censure (p. 116).

It may be that the regressed patients saw the others with whose judgments they differed as potential sources of censure. This hypothesis is strengthened by the fact that more nonconforming errors occurred under the added pressure of the normative and informative conditions where disagreement may have been seen as more readily leading to censure.

On the other hand, it is possible that nonconforming errors represent the *S*s' explicit choice of an alternative which conforms to neither physical nor social reality. It is possible that the *S*s who chose such an alternative did so as a means of insulating themselves from the social environment of the test situation by abstaining from taking action that could be considered relevant by the others present. It can be hypothesized that this abstention functions to inform the others that the *S* cannot be judged according to the rules of the game because he is not a participant. Another hypothesis that views the nonconforming errors as the result of the *S*s' explicit choice of an alternative which conforms to neither physical nor social reality can be made if it is assumed that both agreeing with the majority and going against the majority to give the correct answer are psychologically painful. The nonconforming error may then represent an attempt to abstain from making either commitment. If either of the latter two hypotheses is correct, the posttest errors would represent a continuation of the policy of abstaining from making relevant choices.

A program of research aimed at differentiating among the above hypotheses is now under way. However, the study has direct implications even at present. Both the lack of any significant differences in the amount of conformity to social pressure shown by the *S*s of the different health categories tested and the presence of significant differences in the amount of inappropriate behavior shown by regressed schizophrenics under different conditions of social pressure emphasize the need for a careful re-examination of the nature of schizophrenic withdrawal and its relationship to degree of illness.

SUMMARY

This study was designed to measure the effects of various factors possibly affecting the schizophrenics' readiness to be influenced socially. The basic technique used was the Asch situation. The hypotheses tested were based on the general theory that the schizophrenic process involves a progressive deterioration of responsiveness to the social environment. Thus, it was predicted that the amount of conformity shown would be inversely related to the degree of the subjects' illness.

Two other variables were also investigated. It was hypothesized that schizophrenic subjects to whom the confederates are introduced as normals would conform more than those to whom the confederates are introduced as mentally ill peers. It was also predicted that the amount of conformity exhibited would be increased by the presence of conditions directed at increasing the amount of normative and informative pressure.

Forty-eight regressed subjects, 48 partially remitted subjects, and 48 normal controls were tested. The major hypotheses were not supported by the experimental data. The variables under consideration only affected the amount of conformity shown in terms of a tentatively accepted triple interaction which indicated that for the regressed subjects the informative condition was more effective when presented by normal-confederates than by peer-confederates. In addition, some of the regressed subjects, although able to respond correctly on a pretest, produced a number of responses which conformed to neither the majority's choice nor the correct answer. Furthermore, the number of these responses was increased significantly under the normative and informative conditions, indicating that these regressed subjects are in fact responsive to the immediate social environment.

The results of the study were seen as contradicting the basic hypothesis that the schizophrenic process involves a progressive deterioration of responsiveness to the social environment. Instead of being less responsive than others to the immediate social environment, the regressed subjects were more so, this increased responsiveness, however, taking the form of the production of responses which conform to neither physical nor social reality.

REFERENCES

ASCH, S. D. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**(9, Whole No. 416).

CAMERON, N. *The psychology of behavior disorders.* New York: Houghton Mifflin, 1947.

DEUTSCH, M., & GERARD, H. B. A study of normative and informational influence upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.

FENICHEL, O. The psychoanalytical theory of neurosis. New York: Norton, 1945.

FROMM-REICHMAN, FRIEDA. Transference problems in schizophrenia. *Psychoanal. Quart.*, 1939, **8**, 412–426.

HOVLAND C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.

KRUSKAL, W. H., & WALLIS, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Ass.*, 1952, **47**, 583–621.

RACKOW, L. L., NAPOLI, P. J., KLEBANOFF, S. G., & SCHILLINGER, A. A. A group method for the rapid screening of chronic psychiatric patients. *Amer. J. Psychiat.*, 1952–53, **109**, 561–566.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1957. Pp. 109–184.

SCHOOLER, C. The susceptibility to social influence of chronic schizophrenics in the formation of perceptual judgments. Unpublished doctoral dissertation, New York University, 1959.

SCHOOLER, C., & SPOHN, H. The social structure of a ward of chronic schizophrenics. *Int. J. soc. Psychiat.*, in press.

(Received August 10, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 355–359

DIFFERENCES IN RESPONSE LATENCY AND RESPONSE VARIABILITY BETWEEN HIGH AND LOW ANXIETY SUBJECTS IN A FLICKER-FUSION TASK[1]

ROBERT A. WAGONER[2]
Duke University

THREE firm relationships have been demonstrated between anxiety level and performance in a series of psychophysical tasks. Goldstone (1953, 1955) and Jones (1955, 1958) have shown an inverse relationship between the level of manifest anxiety and measures of flicker-fusion threshold as well as a direct relationship between anxiety and several variability measures. A third index, the number of times a subject changes from one response category to another and back again (alternation of response), was found in a study by Wagoner and Cohen (1956) to be significantly larger for the highly anxious group. This finding was later confirmed by Jones (1955, 1958), who found reliable differences among high, medium, and low anxiety groups on this measure. These alternations may reflect an indecisiveness at the threshold which is exaggerated in the high anxiety group when very fine discriminations are required.

Careful study of some of the response characteristics related to variability around the threshold could shed light on the differences noted between high and low anxiety groups. In the situation confronting the subject in a flicker-fusion task, he is required to make a decision between either the category "flicker" or the category "steady" with regard to any one of the experimental stimuli. If discrete stimuli are presented in a sequence from low to high frequency (or the reverse), sufficient stimulus information is presumably present at the upper and lower extremes to produce consistent use of an appropriate category. In the vicinity of the threshold, the necessity for deciding which of the two available categories is appropriate may create a conflict situation. Cartwright's (1941a, 1941b) finding of heightened latency as a characteristic accompaniment of responses near the threshold suggests another method by which to assess events around the limen.

Typically, one would expect significantly larger response latencies for high anxiety subjects in the vicinity of their threshold than for low anxiety subjects. Unfortunately, this view provides no predictions with respect to the expected response latencies associated with the extremes of the continuum, those areas involving the consistent use of the "flicker" or the "steady" category.

One theoretical position that may help here is that of Taylor (1956) and Spence and Taylor (1951). Their position, derived from the Hullian notion of the interactive effect of drive and habit strength, suggests that a high drive state, such as anxiety, produces facilitation of performance where single or highly dominant response tendencies are present, but produces impairment of performance where multiple competing response tendencies are present. Multiple response tendencies are present in the vicinity of the threshold, and single, highly dominant response tendencies are found where subjects are responding consistently with either the "flicker" or the "steady" category. Consequently, mean latencies of high anxiety subjects should be higher than for low anxiety subjects within the threshold area, but lower mean latencies for the highly anxious subjects should be found where consistent "steady" or "flicker" responses are elicited.

METHOD

Subjects. Twenty high anxiety outpatients (HA) and 20 low anxiety relatives (LA) of outpatients from Duke Hospital, closely matched for age and sex, were selected on the basis of two criteria. For the HA patients, a psychiatric diagnosis of anxiety reaction within a 6-month period was required; for the LA group, the absence of any history of psychiatric difficulties was demanded. In addition, scores of 20 and above for the HA group and of 12 and below for the LA served as cutoffs on the Lorr Anxiety Rating Scale. The Lorr scale was always administered by the experimenter and one of two psychiatrists. Both raters had to agree as to placement above or below the cutoff point, or the potential subject was excluded from the study. Jones (1955, 1958) had previously shown that the Lorr Anxiety Rating Scale correlated .81 with the Taylor Manifest Anxiety Scale and .81 with the Duke Symptom Check List. Such high relationships made it possible to use only the Lorr Anxiety Rating Scale in conjunction with psychiatric diagnosis for the identification of groups.

---

[1] Based on a dissertation submitted to the Department of Psychology of Duke University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Appreciation is gratefully expressed to L. D. Cohen for his supervision of this project.

This investigation was supported in part by Research Grant M887C from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] Now at Veterans Administration Hospital, Indianapolis, Indiana.

Equipment. The apparatus employed in this study was similar to that used by Goldstone (1953, 1955). An electronically activated "glow modulator" tube (Sylvania R1131C) illuminated a 5-mm. circular test patch 30 cm. from the subject's eyes at a 1° visual angle. The stimulus, housed in a black box, appeared as a white light at a brightness of .97 millilamberts in a room dimly illuminated by a 60-w. lamp over the experimenter's recording papers. Binocular foveal fixation was used throughout. Each stimulus, at a light-dark ratio of 60:40 lasted 1.5 seconds. The onset of each stimulus activated a Standard Electric Timer calibrated in 1/100 second intervals. The clock stopped only when the subject either pushed a response lever, fixed in a comfortable armrest position, to indicate a judgment of "flicker" or pulled the lever to indicate "steady."

Procedure. After a partial dark adaptation period of about 5 minutes, each subject viewed a series of flickering lights presented in a modified method-of-limits procedure. A total performance consisted of three sets of trials interrupted by a 30-second rest interval. Each trial consisted of five ascending and five descending runs. Runs were terminated after three successive like responses, and the next run was begun one cycle above (ascending) or one below (descending) the terminal response of the preceding series. The percentage of reports of flicker for each half-cycle frequency for each trial separately and for all trials together were plotted on normal probability paper and a straight line of best visual fit was drawn by the experimenter. That number of cycles associated with the 50% point of flicker responses provided an estimate of the threshold, and the number of cycles difference between the 16% and the 84% points provided an estimate of the variability of response.

RESULTS

A progressive decline in threshold values, independent of anxiety, complicates the calculations of the latency data. Table 1 shows the mean threshold values for the HA and LA groups for Trials I and III. For each group, the difference between the first and last trial is significant by *t* test at the .01 level. Such downward shifts in threshold values during the experiment contraindicated the use of a single threshold for each subject for his entire performance. Rather, it required some procedure whereby the threshold for each run could be utilized. The adjustment adopted was based on

TABLE 1

MEAN THRESHOLD AND VARIABILITY MEASURES FOR 20 HIGH ANXIOUS AND 20 LOW ANXIOUS SUBJECTS

	Lorr Scores Range	Mean Threshold				Mean Variability (all trials)	Mean Alternations of Response
		Mean Age	Total Trials	Trial I	Trial III		
High Anxiety	20–26	33.3	35.7	36.94	35.12	1.12	9.92
Low Anxiety	5–12	34.2	40.1	41.02	39.48	.72	2.47

TABLE 2

HYPOTHETICAL RECORD OF SUBJECT'S RESPONSES TO A VISUAL DISCRIMINATION TASK

Cycles	30		31		32		33		34		35	
Run												
1			0	0	0	0	0	1	1	1		1
2					0	0	0	1	1	1	1	
3			0	0	0	0	0	1	1	1		1
4			0	0	0	1	1	1	1	1	1	
5	0	0	0	0	0	0	1	0	1	1	1	
Total	5	5	5	5	5	4	3	1	0	0	0	0
%0's						80	60	20				

Note.—0 = flicker response; 1 = steady response.

TABLE 3

MEAN RESPONSE LATENCIES IN 1/100 SECONDS OF 20 HIGH ANXIOUS AND 20 LOW ANXIOUS GROUPS FOR THE THREE PORTIONS ON ALL TRIALS

	Mean Latency F (100% Flicker)	Difference (T-F)	Mean Latency T (Threshold)	Difference (T-S)	Mean Latency S (100% Steady)	Difference (S-F)
High Anxiety	107.2	93.0**	200.2	46.3**	153.9	46.7*
Low Anxiety	119.2	43.5*	162.7	36.9*	125.8	6.6
Differences HA-LA	−12.0*		32.5**		38.1*	

* .05 level of confidence.
** .01 level of confidence.

an analysis of single runs, both ascending and descending.

As can be seen in Table 2, any run provides consistent flicker and consistent steady responses and may provide inconsistent alternations of response near the threshold. Thus, three segments of a run can be identified: a flicker portion (F), a steady portion (S), and a middle or threshold portion (T). To ensure an adequate sampling of latencies at the threshold where changes of response category occur, T includes not only latencies associated with alternations of response, which may be absent in some runs, but also those latencies associated with the last flicker response on an ascending run or the last steady response on a descending run. Arraying the data in accordance with the category of response, or portion, frees the response categories from their conventional dependence on the cycles per second at which they occurred and permits greater flexibility in the analysis of the data.

Table 1 shows that the HA group obtained a significantly lower threshold (35.7 cycles) than the LA subjects (40.1 cycles). The 4.4 cycle difference between these means is reliable at better than the .01 level. Mean variability is significantly higher for the HA subjects than for the LA group ($p < .01$). The HA group alternated between response categories more frequently than did the LA group ($p < .01$), by means of Wilcoxon's (1949) nonparametric *t* test. Thus, in contrast to the LA group, the HA subjects generally displayed a reduced sensitivity to the flickering light, dispersed their judgments over a wider range of cycles, and shifted back and forth between response categories more frequently. These data are similar to the previous findings of Goldstone (1953, 1955) and Jones (1955, 1958).

Table 3 presents the mean response latencies associated with each of these portions of the run. Of particular importance is the observation that the HA group requires significantly ($p < .01$) more time to respond in T than does the LA group. Similarly, the HA subjects display a longer mean response latency than the LA subjects in their S ($p < .05$). In F, there is a reversal, and the HA group responds more promptly than the LA subjects ($p < .05$).

Within the LA group, no difference is noted between the mean latencies of their F and S. However, mean latencies of their S and F are both significantly different from those latencies accompanying responses in the T or threshold region ($p < .05$). For the LA subjects, T is associated with higher response latencies in comparison to latencies in their F and S. Like the LA group, the HA subjects require longer to respond in their threshold region than they do in their F ($p < .01$) or S ($p < .05$). Unlike the low anxiety subjects however, HA subjects respond more promptly in their F than in their S ($p < .05$).

DISCUSSION

The latency results clearly indicate that the HA group requires more time to respond to threshold stimuli than do the LA subjects. These results support the general hypothesis that increased variability of response in the HA group is associated with increased difficulty in making prompt and effective decisions at the threshold. While elevated relative to their own nonthreshold latencies, mean times at the threshold for the LA subjects did not reach the length of delay characteristic of the

HA group. Thus, the generally longer latencies of the HA group may indicate an exaggeration of a process found also in the LA group. Alternatively, these data may be explained with equal plausibility in terms of competing response tendencies. If mutually exclusive responses separately approach a maximal probability of occurrence at the threshold, increased time may be necessary for the most dominant one to emerge behaviorally in LA subjects. The presence of increased drive in the HA group could elevate the latencies, as observed in this experiment, above the level typical for the LA group.

Advancing the drive conception of anxiety as an alternative explanation of events occurring at the threshold presupposes the adequacy of the theory in dealing with responses not in competition. Since S and F consist of clearly dominant responses, the facilitative effects of heightened drive should reduce the mean of the HA group in these below that of the LA group. This decrement was observed only in F, where both groups were consistently using the flicker category. Contrary to expectation, the HA group took longer to respond in S than did the LA group. This reversal casts some doubt on the applicability of the drive conception of anxiety as applied to this experimental situation.

In this connection, one should point out that the task confronting the high and the low anxiety subjects is for each group somewhat different. Figure 1 shows the mean latencies of response arrayed around the common fixed threshold points (T). This fixed threshold, which was originally calculated to make it independent of cps, is plotted here to fall at the same number of cps as the conventional thresholds for the HA and LA groups. Only in the limited range of cycles falling between 35.0 and 40.0 were the two groups viewing a common series of stimuli. While the discrepancy in the exposed stimulus series perhaps did not interfere with the attempt to clarify events occurring at the threshold, different stimulus situations confronting the two groups may have contributed to some of the differences noted at the extremes of the stimulus series.

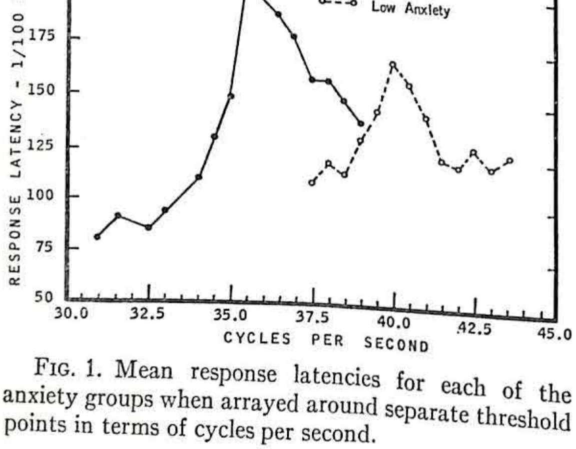

FIG. 1. Mean response latencies for each of the anxiety groups when arrayed around separate threshold points in terms of cycles per second.

The stimuli falling in the range between 35.0 and 40 cps are frequently noted to display a faint shimmering character (Gebhard & Mowbray, 1955). It is possible that the HA and LA groups adopt different criteria to deal with stimuli falling in this range. Different criteria may produce, on the one hand, sharply differentiated thresholds, while, on the other, they make it very difficult for the HA subjects to decide whether or not the stimuli in this range are steady or just faintly shimmering. If this is an adequate characterization of the stimulus situation confronting the HA group, sheer ambiguity of the stimulus series in this range may result in longer latencies. It is of some importance that, whatever these criteria may be, they are self-adopted and appear to be related to the variable of anxiety.

SUMMARY

On the assumption that a subject's decision to use either of two available categories of response at the threshold in a flicker-fusion task is much like a conflict situation, HA subjects were expected to take more time to resolve the conflict and hence, display longer latencies at threshold than LA subjects. Where consistent responses were possible outside the threshold region, the drive conception of anxiety would predict that the HA subjects respond more promptly than LA subjects. These two hypotheses were tested in a flicker-fusion situation where latencies were recorded for each response. Twenty HA psychiatric outpatients and 20 LA relatives were matched for age and sex. The mean threshold latency of the HA group was considerably larger than that of the LA group, although the threshold latency of

each group was elevated above its own nonthreshold latency. Latencies accompanying nonthreshold responses were lower for the HA group only where subjects consistently responded with the category "flicker." Where consistent "steady" responses were possible, the LA group displayed shorter mean latencies than the HA group, a finding contrary to the hypothesis based on the drive conception of anxiety. This latter result may be a function of either the inappropriateness of the drive conception of anxiety to this task or an inherent stimulus ambiguity.

REFERENCES

CARTWRIGHT, D. Decision-time in relation to differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442. (a)

CARTWRIGHT, D. The relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 177–196. (b)

GEBHARD, J. W., & MOWBRAY, G. H. Differential sensitivity of the eye to intermittent white light. *Science*, 1955, **121**, 173–175.

GOLDSTONE, S. Flicker fusion measurements and anxiety level. Unpublished PhD thesis, Duke University, 1953.

GOLDSTONE, S. Flicker fusion measurements and anxiety level. *J. exp. Psychol.*, 1955, **49**, 200–202.

JONES, O. Relationship between visual and auditory discrimination and anxiety level. Unpublished PhD thesis, Duke University, 1955.

JONES, O. Relationship between visual and auditory discrimination and anxiety level. *J. gen. Psychol.*, 1958, **59**, 111–118.

SPENCE, K., & TAYLOR, JANET. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

TAYLOR, JANET. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

WAGONER, R., & COHEN, L. D. Analysis of patterns of response of anxious and non-anxious subjects to a flickering light. *Percept. mot. Skills*, 1956, **6**, 167–170.

WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Co., 1949.

(Received July 20, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 360–364

LEADER-MEMBER INTERACTION IN MANAGEMENT COMMITTEES

LAWRENCE SCHLESINGER,[1] JAY M. JACKSON,[2] AND JEAN BUTMAN
University of Michigan

STUDIES in leadership have shifted from a concern with characteristics and traits of leaders to a concern with the process of leadership, and with the exertion and acceptance of influence. As one step in this direction, the behavior of leaders has been reliably differentiated from that of members in analyses of the interaction processes in discussion groups. Kirscht, Lodahl, and Haire (1959) have recently summarized the existing data on categorical differences:

> . . . the general picture of a small group leader which appears common to various researches shows a group member who is . . . task-oriented, attempts to specify the problem, to suggest courses of action, to seek out the members' contributions, to integrate these and to propose solutions in the attempt to secure consensus in the group (p. 408).

At least one dimension, then, in which leader behavior differs from that of members may be described as the attempts of the leader to exert influence or control over the behavior of the members. Leadership acts may differ in the magnitude of attempted control as well as in frequencies of attempted control acts. The differences reported in frequency of attempted control behavior by perceived or designated leaders leaves open the question of the process whereby the behavior of the leader comes to be differentiated from that of the members. Do leaders of small groups actually make more influence attempts than the other members? Or do group members diminish their attempts to control the behavior of their fellow members?

Some evidence on these questions comes from a study in which the degree of control by the chairmen of existing management committees was manipulated by experimental appointment of chairmen who, in the opinion of their fellow members, varied in their directiveness. This experiment provided an opportunity to observe the influence process between chairmen and members under chairmen who differed in the degree of control they were perceived as exerting over group members.

METHOD

Subjects. The study was conducted in a large utility company, on 23 existing committees of 106 middle-management executives. Each committee contained four or five members representing a major operating department and had the responsibility of coordinating interdepartmental problems in geographical districts of the company. The larger study[3] was designed to evaluate conditions influencing the performance of these interdepartmental management committees.

Selection of experimental chairmen. The independent variable in this study is the degree of control exercised by the designated chairmen of the committees. This manipulation was achieved without the knowledge of any of the participants including the chairman himself. Two months prior to the field experiment, all members of the committees responded to an extensive mailed questionnaire including a Chairman Behavior Questionnaire. In this questionnaire, they were asked to describe the typical behavior of each committee member when he was in the role of chairman, a position which rotated regularly among the members. The items used to elicit this description stem from the definition of "control" as a set of behaviors aimed at inducing restrictions on the behavior of the recipients. Scores derived from this questionnaire indicate the frequency with which the committee member in the role of chairman acts in the following way, according to other members of his committee: he does a great deal of guidance, he makes proposals for handling problems, he stimulates the group to action, he steers the discussion, he frequently gives his own opinions, and he often questions the statements of the members.

Using the mean of scores given by all other members of his group, then, a person was designated as a High Control or Low Control chairman depending upon whether his score fell above or below the median, both overall and in his own committee. These mean scores ranged, on a nine-point scale, from 7.56 to 2.17 with the median score at 5.04. We attempted to have the same number of departmental and divisional affiliations in each leadership condition.

Procedure. The team of investigators traveled to 23 different sites, where the interdepartmental committees usually met, but standardized all procedures and collected comparable data from each committee. Each committee worked on the solution of a simulated but realistic problem for 1.5 hours under controlled conditions in the presence of three observers. The problem given to each committee was complex, involving considerable detailed information. It had been constructed by the investigators, with the help of a committee of executives, from problems typical of those faced by the committees in their regular meetings.

---

[1] Now at 708 West Huron Street, Ann Arbor, Michigan.

[2] Now at University of Kansas.

[3] Institute for Social Research, Research Center for Group Dynamics. A study of Interdepartmental Management Committees. Unpublished manuscript, Ann Arbor, 1958.

Each member was provided with an information sheet relevant to his department at the beginning of the meeting. The chairman was designated by the principal investigator, supposedly on a chance basis, and provided with additional instructions. The latter required that each group produce a detailed solution including recommended actions to implement their decision. The chairman was instructed to write up the minutes of the meeting, summarizing the decisions and recommendations reached during the discussion. The interest of higher management in the conduct and outcome of the study had been communicated to the executives, increasing the concern of the committees with their performance. Emphasis, however, was placed on the performance of the committee as a whole, and the executives were explicitly told that the study was not an assessment of individual abilities.

Data collected. The data relevant to this paper include observing and recording interactions during the meeting. Two management trainees of the company assigned to the study to act as observers recorded the interactions in each meeting with a system of categories based on Bales' (1950) system. The three relevant categories in the "control" dimension vary in magnitude of attempted control.

Control—statements aimed at restricting the behavior of the recipients. Three degrees of control are distinguished, depending on the latitude for compliance allowed the recipient:

1. Direction—the speaker orders or directs the other person or persons to whom the remark is addressed to do something giving the other little opportunity to refuse, e.g.: "We should read these first, then start the discussion."
2. Suggestion—the speaker suggests that the other do something, giving him room to refuse or make a countersuggestion, e.g.: "I suggest that Jack might start off by describing the situation in his department."
3. Request—the speaker asks the other to do something implying he has little right to ask and the other has every right to refuse, e.g.: "Would you be willing to tackle the problem of calculating the cost of the second alternative, or do you feel that can wait until we see the whole picture?"

RESULTS

Participant pre- and post-meeting ratings. The original ratings of how the person behaved as chairman and ratings on the same items immediately after the experimental meeting, with that person as chairman, correlated .72, significant well beyond the .01 level for the 23 groups. The chairmen selected for the experimental conditions were seen to exhibit the same degree of control in the experimental meetings that the members had described in the questionnaire several months earlier.

Participant and observer ratings. The chairman's degree of control, as assessed by three neutral observers, was combined into a pooled rating that correlated .60 with the combined rating by participants, again significant well beyond the .01 level.

Interactions of experimental chairmen. Consistency of interaction recording between the two observers was measured by correlating the frequencies of interactions in a category attributed to a person during a meeting by each of the observers. The analysis was performed for a sample of five meetings. Where no interaction was attributed in a category by either observer, that pair of zero values was not included in the computation of the correlation, a decision which depressed the value of the correlations. As Table 1 indicates, the reliabilities thus measured show considerable variability, with only the "rejects" category reaching a less than acceptable level.

In describing the behavior of the chairman, two scores are possible: scores that describe the chairman's performance in relation to self, and ones that describe his performance in relation to the total group (Heyns & Lippitt, 1954). For each chairman there are two scores, one that expresses the concentration of verbal behavior in each category as a percentage of the chairman's own total behavior, and one that indicates what part of the total participation by the group, in each category, is accounted for by the chairman.

TABLE 1

CORRELATIONS BY INTERACTION CATEGORY BETWEEN OBSERVERS IN FREQUENCY OF TALLIES ATTRIBUTED TO SAMPLE OF INDIVIDUALS

	Categories of Interaction						
	Directs	Suggests	Requests	Accepts	Rejects	Gives Orientation	Asks for Orientation
r	.57*	.91**	.55*	.83**	.38	.51*	.95**
N	17	22	19	28	26	28	24

* Significant at .05 level.

** Significant at .01 level.

Looking first at the chairmen's specific acts in comparison to their own total behavior, in Table 2, there is no evidence that a chairman's degree of control score is significantly related to the percentage of his own behavior falling in any of the categories of interaction. Although the chairmen were reliably rated as exerting different degrees of control by both participants and observers, neither the magnitude nor frequency of control acts observed was related to rated degree of control. Since the percentage distributions may conceal the fact that High Control chairmen are more active in all categories than Low Control chairmen, a comparison was made of the absolute amount of activity of each of the two types of chairmen. The results were substantially the same, with no difference in the distributions or level of observed interactions. However, in Table 2, viewing the chairman's behavior in comparison with the other members of his group, i.e., as the percentage of the group's acts in various categories, the distinction between types of chairmen becomes clearer. In comparison to other members of their committee, High Control chairmen performed many more acts in the control dimension than Low Control chairmen, with highly significant correlations in both the suggests and requests categories. Viewing the behavior of the members alone, in the last column of Table 2, we find a negative relation between the percentage of the group's acts that are directive and the degree of control exercised by the chairman. Only the "requests" category reached significance.

TABLE 2
RELATIONSHIPS BETWEEN CHAIRMAN'S DEGREE OF CONTROL AND CONTROL BEHAVIOR OBSERVED IN EXPERIMENTAL MEETING

Category of Interaction	Correlations[a] with Chairman Control		
	Chairman's Behavior		Member Behavior
	Percentage of Chairman's Acts	Percentage of Group's Acts	Percentage of Group's Acts
Directs	.16	.28	−.32
Suggests	.15	.51**	−.16
Requests	.19	.62**	−.41*

[a] Pearson product-moment.
* Significant at the .05 level.
** Significant at the .01 level.

TABLE 3
RELATIONSHIPS BETWEEN MEMBER RATINGS ON DEGREE OF CONTROL EXERCISED AS CHAIRMEN AND OTHER CHARACTERISTICS

Member Characteristics	Correlation[a]
Age	−.29
Length of service	−.33
Boss' influence on him	−.26
His influence on boss	.09
Ratings by others:	
Contribution to meetings	.56*
Skill as chairman	.75*
Friendly relations	.14

[a] Pearson product-moment.
* Significant at the .01 level.

Differences among experimental chairmen. To ascertain the significance to participants of the score for degree of control, the control score obtained from the Chairman Behavior Questionnaire was related to a number of other characteristics of the chairmen, and to ratings of them by the participants, obtained in a questionnaire filled out prior to the experimental sessions. The analysis was based on the entire sample of 106 members of the management committees.

In Table 3 it can be seen that High Control chairmen show some tendency, not statistically significant, to be younger, to have less service with the company and to feel somewhat freer of influence from their immediate superiors. The findings that reach a high level of significance refer to the way committee members rate each other on their skill as chairmen and on their general contribution to committee meetings. Members who are seen as high on control are rated as more skillful chairmen and also as contributing more to committee proceedings in their regular membership role. No relation was found between degree of control and feelings of personal friendliness among members. The results were substantially the same when the analysis was based on the 23 appointed chairmen.

Discussion

The problem of this paper is to account for the differences in interaction patterns of leaders and group members that have been reported in the literature. Conceptually, the elements of this experiment may be considered the goal of the committee members, the role of the chairman, the magnitude of control exerted by the chairman over the members, the frequency of these attempted control acts by chairman and members, and the perceived resources of the chairman. First it seems reasonable to assume that as a result of the experimental instructions, and in the context of an evaluation of the committees, the goal of the committees was to provide a high quality solution to the problem presented. The role of the designated chairman may be conceived as a legitimate power role, following French and Snyder (1959) in which the incumbent has a legitimate right to influence the members and they an obligation to accept his influence. The two chairmen's interaction scores, based on the percentage of their own acts and the percentage of group acts in a category will be referred to as the individual and category scores, respectively, and have different conceptual meanings (Heyns, 1948). The individual score indicates the magnitude of control the person attempts to exert in response to the demands of the chairman role. The category score reflects the role the person plays in the group, including both his response to role demands and the acceptance of his influence by the members.

Considering first the chairmen's individual roles, the different chairmen evidently responded similarly to the demands on them in their role to direct and guide the meeting. Although both the participants and observers rated their behavior differently on the control dimension, they did not actually differ in the magnitude or frequency of control acts.

The category scores indicate that the High and Low Control chairmen had considerably different roles in the group. Compared to other group members, High Control chairmen exerted much more influence than did Low Control chairmen. Similar findings have been reported in Heyns' (1948) laboratory study of leader-member interaction patterns under chairmen instructed to either exercise or fail to exercise directive and guidance functions. Inspection of his data, which used identical scoring procedures and similar categories, indicates the greatest observed differences in controlling behaviors among leaders when their interactions are seen in relation to member behavior. The notion that leadership is not an attribute of a single agent, but a relationship among persons is strikingly supported.

The High Control chairmen, in their category scores, actually exerted a lesser magnitude of control, than did the Low Control chairmen. Their guidance implied relative autonomy for the members in comparison to directive control which leaves the members little choice but outright rejection or compliance. This combination of a higher frequency and lesser magnitude of control by the High Control chairmen suggests differences in response of the participants to the control acts of the chairmen.

When the participants do not perceive the designated leader as satisfactorily performing the controlling functions, the participants increase their own attempts to influence their fellow members. Heyns (1948) also reported that participants under the leader who was less controlling tended to perform more of the guidance functions than participants under the chairman who did exercise control. Carter, Haythorn, Shriver, and Lanzetta (1951) also reported that participants exerted more influence when the designated leaders did not respond directively. The greater frequency of influence attempts by participants in the Low Control groups did not crystallize into a greater emergence of informal leaders in these groups, as some industrial studies have reported. When the designated leader fails to perform adequately, considerable time is probably necessary for the members to decide which alternative persons they can confer leadership status upon.

The question remains of accounting for differences in participant response to the two kinds of leaders. Theoretically, leadership acts that are perceived as enabling the group members to reach their goal would be most acceptable to the participants. There is little doubt that the High Control chairmen were perceived as most capable of performing these leadership acts. The finding that degree of control was positively associated with two kinds of expertness, contributor to the group's tasks and skill as chairman, parallels Bales' (1953) report that the "best ideas" and "guidance" roles are closely associated in the participants' minds with leadership. As French and Snyder (1959) have reasserted, the effectiveness of an influence attempt by the leader increases with increasing perception that he is an expert in the area of the influence attempt.

It is noteworthy that leader acceptance, in this experiment, was based on task and chairmanship skills, and completely distinguished, as Bales (1953) found, from relations of interpersonal liking. As Gibb (1954) has suggested, the leader's influence upon group members is secondary to his influence upon goal locomotion. Although each of the chairmen attempted to exert the same amount of influence, the group members conferred leadership status upon those chairmen who were seen as providing a pathway to the group goal.

SUMMARY

The problem of this study was to examine the influence process among leaders and members of small problem solving groups when the designated leaders vary in rated degree of control exerted. The opportunity for the study was provided by a field experiment on the degree of control exercised by the designated chairmen of management interdepartmental committees.

As part of an initial questionnaire, 106 members of 23 interdepartmental middle-management committees described, on a six-item questionnaire, the amount of control and regulation each member exercised in the role of chairman. A special meeting was then held in the field with half the committees under High Control, and half under Low Control chairmen, appointed "randomly" by the investigators. Each committee was observed for 1.5 hours while working on a simulated but typical company problem. The goal of the committees was to present the best possible solution of this problem to be assessed by higher management.

Observer and participant ratings of the chairman's behavior validated the original questionnaire ratings. But interaction profiles, based on the number of acts by a chairman in each category, showed no difference between the types of chairmen. They did differ strikingly from one another, however, when interaction scores were computed as ratios of the total number of acts performed by the entire committee in a particular category. Compared to other committee members, High Control chairmen were higher on the percentage of "control" acts performed. Committee members under Low Control chairmen made more attempts to influence group members. Analysis of the characteristics associated with the perceived degree of control exercised by the member when he was chairman indicated that the High Control chairman was more likely to be considered a skilled chairman and a more valuable contributor to the committee's work than the Low Control chairman.

It was concluded that the observed difference in leader-member interaction patterns is conditioned by the acceptance of the designated chairman as a resource by the group members. When the designated chairman is accepted as a resource capable of directing the group toward a successful solution of their task, he exerts a more frequent and lesser magnitude of control. The influence attempts of the Low Control leader are not accepted by group members, who vie with one another and the designated chairman for control over the group's activities. Differences in leader-member interaction patterns are thus seen to result from differences in the response of the members to leadership acts which vary in their perceived capacity to lead members to their goal.

REFERENCES

BALES, R. F. *Interaction process analysis.* Cambridge: Addison-Wesley, 1950.

BALES, R. F. The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. A. Shils (Eds.), *Working papers in the theory of action.* Glencoe, Ill.: Free Press, 1953. Pp. 111–161.

CARTER, L., HAYTHORN, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589–595.

FRENCH, J. R. P., JR., & SNYDER, R. Leadership and interpersonal power. In D. Cartwright (Ed.), *Studies in social power.* Ann Arbor: Institute for Social Research, 1959. Pp. 118–149.

GIBB, C. Leadership. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. II. Cambridge: Addison-Wesley, 1954. Pp. 877–920.

HEYNS, R. W. Effects of variation in leadership on participant behavior in discussion groups. Unpublished doctoral dissertation. Ann Arbor: University of Michigan, 1948.

HEYNS, R. W., & LIPPITT, R. Systematic observation techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge: Addison-Wesley, 1954. Pp. 370–404.

KIRSCHT, J. P., LODAHL, T. M., & HAIRE, M. Some factors in the selection of leaders by members of small groups. *J. abnorm. soc. Psychol.*, 1959, **58**, 406–408.

(Received August 11, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 365-369

COMPETENCE AND CONFORMITY IN THE ACCEPTANCE OF INFLUENCE[1]

E. P. HOLLANDER

School of International Service, American University

When one member influences others in his group it is often because he is competent in a focal group activity. A member may show such competence by individual actions that further the attainment of group goals (cf. Carter, 1954); more specific situational demands may variously favor the ascent of the expediter, advocate, or what Bales and Slater (1955) have termed the task specialist. An additional condition for the acceptance of influence involves the member's perceived adherence to the normative behaviors and attitudes of his group. His record of conformity to these expectancies serves to sustain eligibility of the sort Brown (1936) calls "membership character."

A person who exhibits both competence and conformity should eventually reach a threshold at which it becomes appropriate in the eyes of others for him to assert influence; and insofar as these assertions are accepted he emerges as a leader. But it is still necessary to account for the "nonconformity" that leaders display as they innovate and alter group norms. Certain shifts must therefore occur in the expectancies applicable to an individual as he proceeds from gaining status to maintaining it.

This process has been considered recently in a theoretical model of status emergence (Hollander, 1958). It features the prospect that behavior perceived to be nonconformity for one member may not be so perceived for another. Such differentiations are seen to be made as a function of status, conceived as an accumulation of positively disposed impressions termed "idiosyncrasy credits." A person gains credits, i.e., rises in status, by showing competence and by conforming to the expectancies applicable to him at the time. Eventually his credits allow him to nonconform with greater impunity.[2] Moreover, he is then subject to a new set of expectancies which direct the assertion of influence. Thus, whether for lack of motivation or misperception, his failure to take innovative action may cause him to lose status.[3]

It is readily predictable that in task oriented groups a member giving evidence of competence on the group task should with time gain in influence. If he simply nonconforms to the procedures agreed upon, the opposite effect should be observed. But the sequential relationship of nonconformity to competence is especially critical.

From the model, it should follow that, with a relatively constant level of manifest competence, the influence of a person who nonconforms *early* in the course of group interaction should be more drastically curtailed than in the case of a person who nonconforms *later*. Indeed, a reversal of effect would be predicted in the latter instance. Once a member has accumulated credits, his nonconformity to general procedure should serve as a confirming or signalizing feature of his status, thereby enhancing his influence. Accordingly, it may be hypothesized that given equivalent degrees of task competence, a member should achieve greater acceptance of his influence when he has conformed in the past and is now nonconforming than he should when nonconformity precedes conformity.

METHOD

Design

Twelve groups, each composed of four male subjects, were engaged in a task involving a sequence of

---

[1] This paper is based upon a study completed under ONR Contract 1849(00), while the author was at the Carnegie Institute of Technology. The views expressed here are those of the author and do not necessarily reflect those of the Department of the Navy.

The considerable assistance of H. Edwin Titus in this study is gratefully acknowledged.

Parts of the paper were reported at the symposium on "Recent Conceptions in Influence and Authority Process," held under the auspices of Division 8 at the 1959 APA Convention.

[2] This is a newer formulation of an observation long since made regarding the latitude provided leaders (e.g., Homans, 1950, p. 416). It is further elaborated in Hollander (1959).

[3] This proposition is consistent with various findings suggestive of the greater social perceptiveness of leaders (e.g., Chowdhry & Newcomb, 1952).

	Green	Red	Blue	Yellow	Brown	Orange	Black
Able	−1	−12	+5	−1	−2	+15	−4
Baker	+10	−1	−2	−7	+4	−3	−1
Charlie	−5	+5	−3	+3	−11	−1	+12
Dog	+5	−7	+10	−2	−5	+1	−2
Easy	−4	−1	−1	+1	+13	−10	+2
Fox	−6	+15	−5	−1	−3	−1	+1
George	−1	−1	−2	+10	+4	−2	−8

FIG. 1. Matrix used in group task.

15 trials. A group choice was required for each trial from among the row alternatives in a 7 × 7 payoff matrix (see Figure 1). In every group, a fifth member was a confederate whose prearranged response was contrived to be correct on all but four trials, i.e., 2, 3, 6, and 12, thus reflecting considerable competence on the task. All interactions among participants took place through a system of microphones and headsets from partitioned booths. Subjects were assigned numbers from 1 to 5 for communicating with one another. The central manipulation was the confederate's nonconformity to procedures agreed upon by each group in a pretrial discussion. In terms of a division of the 15 trials into three zones—early, middle, and late—of 5 trials each, six treatments were applied: nonconformity throughout, nonconformity for the first two zones, for the first zone alone, for the last two zones, for the last zone alone, and a control with no nonconformity. In one set of treatments the confederate was designated number 5, and in the other number 4, to test possible position effects. Acceptance of the confederate's influence was measured by the number of trials by zone in which his recommended response was accepted as the group's. This was supplemented by post-interaction assessments.

Subjects

The 48 subjects were all juniors in the College of Engineering and Science at the Carnegie Institute of Technology. All had volunteered from introductory psychology sections after being told only that they would be taking part in a study of problem solving in groups. Care was taken in composing the 12 groups so as to avoid either placing acquaintances together or having membership known in advance. Thus, no two subjects from the same class section were used in the same group, and subjects reported at staggered times to different rooms. By the time a subject reached the laboratory room where the experiment was actually conducted, he had been kept apart from the others and was not aware of their identity. The subjects never saw one another during the entire procedure, nor were their names ever used among them.

Instructions and Set

Once seated and assigned a number, every subject was given a sheet of instructions and the matrix used for the task. These instructions fell into two parts, both of which were reviewed aloud with each subject individually, and then with the entire group over the communication network. The first part cautioned the subjects to always identify themselves by number (e.g., "This is Station 3 . . .") before speaking and not to use names or other self-identifying references. The second part acquainted them with the procedures to be used, emphasized the aspect of competition against a "system," and established the basis for evident procedural norms. It read as follows:

1. You will be working with others on a problem involving a matrix of plus and minus values. Everyone has the same matrix before him. The goal is to amass as many plus units as possible, and to avoid minus units. Units are worth 1 cent each to the group; the group begins with a credit of 200 units. You cannot lose your own money, therefore. There will be fifteen trials in all.

2. In any one trial, the task involved is for the group to agree on just *one* row—identified by Able, Baker, Charlie, etc.—which seems to have strategic value. Once the group has determined a row, the experimenter will announce the column color which comes up on that trial. The intersecting cells indicate the payoff. Following this announcement, there will be thirty seconds of silence during which group members can think individually about the best strategy for the next trial, in terms of their notion about the system; note please that there are several approximations to the system, although the equation underlying it is quite complex. But work at it.

3. At the beginning of each trial the group members must report, one at a time, in some order, as to what they think would be the best row choice on the upcoming trial. Members may "pass" until the third time around, but must announce a choice then. Following this, groups will have three minutes on each trial to discuss choices and reach some agreement; this can be a simple majority, or unanimous decision; it is up to the group to decide. If a decision is not reached in three minutes, the group loses 5 units.

4. Before beginning the trials, the group will have five minutes to discuss these points: (*a*) The order of reporting; (*b*) How to determine the group choice for a given trial; (*c*) How to divide up the money at the end. These decisions are always subject to change, if the group has time and can agree. After the 15th trial, group members may have as much as five minutes to settle any outstanding decisions. Then headsets are to be removed, but group members remain seated for further instructions, and the individual payment of funds.

Instruments and Procedure

The matrix was specially constructed for this study to present an ambiguous but plausible task in which alternatives were only marginally discrete from one another.[4] The number of columns and rows was selected to enlarge the range of possibilities beyond the number of group members, while still retaining comprehensibility. The fact that the rows are unequal in algebraic sum appears to be less important as a feature in choice than the number and magnitude of positive and negative values in each; there is moreover the complicating feature of processing the outcome of the last trials in evaluating the choice for the next. All considered, the matrix was admirably suited to the requirements for ambiguity, challenge, conflict, immediate reinforcement, and ready manipulation by the experimenter.

The confederate, operating as either 4 or 5 in the groups, suggested a choice that differed trial by trial from those offered by other members; this was prearranged but subject to modification as required. Since subjects rather typically perceived alternatives differently, his behavior was not unusual, especially during the early trials. For the 11 trials in which the confederate's row choice was "correct," the color that "came up" was contrived to yield a high plus value without at the same time providing a similar value for intersection with another person's row choice. Had his recommendation been followed by the group on these trials, high payoffs would have accrued.

The device of a 5-minute pretrial discussion had special utility for establishing common group expectancies, in the form of procedures, from which the confederate could deviate when called for in the design. Predictable decisions on these matters were reached unfailingly. But their importance lay in having a *public affirmation* of member intent. Thus, on order of reporting, it was quickly agreed to follow the order of the numbers assigned members. Each group, despite minor variants suggested, decided on simple majority rule. Regarding division of funds, equal sharing prevailed, sometimes with the proviso that the issue be taken up again at the end.

In the zones calling for nonconformity, the confederate violated these procedures by speaking out of prescribed turn, by questioning the utility of majority rule, and by unsupported—but not harsh—challenges to the recommendations made by others. He manifested such behaviors on an approximate frequency of at least one of these per trial with a mean of two per trial considered optimum. Thus, he would break in with his choice immediately after an earlier respondent had spoken and before the next in sequence could do so; when there were periods of silence during a trial he would observe aloud that maybe majority rule did not work so well; and he would show a lack of enthusiasm for the choice offered by various others on the matter of basis. Lest he lose credibility and become a caricature, in all instances he chose his moments with care and retained an evident spontaneity of expression.[5]

RESULTS AND DISCUSSION

The task gave quite satisfactory signs of engrossing the subjects. There was much talk about the "system" and a good deal of delving into its basis, possibly made the more so by the subjects' academic background; the returned matrices were littered with diagrams, notations, and calculations. Though quite meaningless in fact, the confederate's tentative accounts of his "reasoning" were evidently treated with seriousness, perhaps as much because of the contrived time constraint, which prevented probing, as of his jargon regarding "rotations" and "block shifts." In any case, the confederate at no time claimed to have the system completely in hand. He delayed his response from the sixth trial onward to suggest calculation of an optimum choice in the face of conflicting alternatives; and the four trials on which he was "wrong" were spaced to signify progressive improvement, but not total perfection.

Most pertinent, however, is the fact that there were no manifestations of suspicion concerning the confederate's authenticity. The others seemed to believe that he was one of them and that he was "cracking" the system; the post-interaction data were in full agreement.

Since all of the interactions were available on tape, it was possible to derive a number of indices of acceptance of influence. The most broadly revealing of these appeared to be the frequency of trials on which the confederate's recommended solution was followed.

In Table 1 this index is employed to discern the effects of three major variables. The analysis is arranged by zones (Z) of trials, and in terms of the confederate's nonconformity (NC) in the *current* zone and immediate *past* zone.[6] The means given in each cell indicate the number of trials, out of five per zone, on which the confederate's choice was also the group's. In a chi square test, the effect of position upon this measure was found to be nonsignificant, and is therefore omitted as a distinction in the analysis of variance.

---

[4] The matrix is an adaptation, at least in spirit, of a smaller one used with success by Moore and Berkowitz (1956).

[5] The same person, H. E. Titus, was the confederate throughout.

[6] For Zone I, the "past zone" refers to the discussion period. If he was to nonconform there, the confederate would question majority rule and suggest that the division of funds be left until the end rather than agree then on equal shares.

TABLE 1

MEAN NUMBER OF TRIALS ON WHICH A GROUP ACCEPTS CONFEDERATE'S RECOMMENDED SOLUTION

Confederate's Previous Conformity	Zone I (Trials 1–5)		Zone II (Trials 6–10)		Zone III (Trials 11–15)	
	Nonconforming[a]	Conforming	Nonconforming	Conforming	Nonconforming	Conforming
With Procedural nonconformity in immediate *past* zone	1.67 6[b]	—	3.25 4	3.00 2	4.00 4	5.00 2
Without Procedural nonconformity in immediate *past* zone	—	2.00 6	5.00 2	3.75 4	5.00 2	4.75 4

ANALYSIS OF VARIANCE

Source	*SS*	*df*	*MS*	*F*
Current Nonconformity	.20	1	.200	—
Zones	47.05	2	23.525	35.01**
Past Nonconformity	3.36	1	3.360	5.00*
Int: Current NC × Z	1.22	2	.610	—
Int: Current NC × Past NC	13.52	1	13.520	20.12**
Int: Z × Past NC	.72	2	.360	—
Int: Current NC × Z × Past NC	4.11	2	2.055	3.06
Residual	16.12	24	.672	
Total	86.30	35		

[a] Confederate showed procedural nonconformity on the trials in this zone.

[b] Indicates number of groups upon which cell is based.

* $p < .05$.

** $p < .001$.

The significant *F* secured from Zones is in accord with prediction. It reveals the ongoing effect of task competence in increasing the acceptance of the confederate's choice, to be seen in the rising means across zones. While current nonconformity does not yield a significant effect, past nonconformity does. Viewing the table horizontally, one finds that the means for "without" *past* NC exceed the means for "with" *past* NC in all instances but one. Regarding the significant interaction of *current* and *past* NC, the combination "without-without" has a sequence (2.00, 3.75, 4.75) of persistently higher value than has "with-with" (1.67, 3.25, 4.00); this, too, is in line with prediction. Finally, the maximum value of 5.00 in Zone II for the combination "without" *past* NC but "with" *current* NC confirms the key prediction from the model, at least within the context of the relative magnitudes there; the same value is also seen in Zone III for the identical combination; still another reading of 5.00 holds there, however, for the inverse combination, but in a tight range of values quite beyond separation of effects for interpretation.

Considerable consistency was found too in the post-interaction data. On the item "overall contribution to the group activity," 44 of the 48 subjects ranked the confederate first; on the item "influence over the group's decisions," 45 of the 48 ranked him first. Two things bear emphasis in this regard: subjects had to individually write in the numbers of group members next to rank, hence demanding recall; and their polarity of response cut across all six treatments, despite significant differences among these in the actual *acceptance of influence*. That the confederate therefore made an impact is clear; but that it had selective consequences depending upon the timing of his nonconformity is equally clear.

In detail, then, the findings are in keeping with the predictions made from the model. The operational variable for measuring acceptance of influence was confined to the task itself, but nontask elements are touched as well. In that respect, the findings corroborate the subtle development of differential impressions as a function of even limited interpersonal behavior.

Some unquantified but clearly suggestive data are worth mentioning in this regard. Where, for example, the confederate began nonconforming *after* the first zone, his behavior was accepted with minimal challenge; by the third zone, his suggestion that majority rule was faulty yielded a rubber stamping of his choice. Again, if he had already accrued credit, his pattern of interrupting people out of turn not only went unhindered but was taken up by some others. Quite different effects were elicited if the confederate exhibited nonconformity from *the outset*, notably such comments of censure as "That's not the way we agreed to do it, five."

The findings are especially indicative of the stochastic element of social interaction and its consequence for changing perception. Especially interesting is the fact that these effects are produced even in a relatively brief span of time.

SUMMARY

A study was conducted to test the relationship between competence on a group task and conformity or nonconformity to procedural norms in determining a person's ability to influence other group members. Data were gathered from 12 groups engaged in a problem solving task under controlled conditions. Each was made up of five members one of whom was a confederate who evidenced a high degree of competence during the 15 trials. His nonconformity to the procedural norms agreed upon by the group was introduced at various times, early, middle, or late, in the sequence of trials. Influence was measured by the number of trials (per segment of the entire sequence) in which the confederate's recommended solution was accepted as the group's choice. As a broad effect, it was found that a significant increase in his influence occurred as the trials progressed, presumably as a function of the successive evidences of competence. Past conformity by the confederate was also found to be positively and significantly related to the acceptance of his influence; finally, there was a statistically significant interaction between past and current nonconformity reflected in high influence in the groups in which the confederate had conformed earlier in the sequence of trials but was presently nonconforming. These results were all thoroughly consistent with predictions made from the "idiosyncrasy credit" model of conformity and status.

REFERENCES

BALES, R. F., & SLATER, P. E. Role differentiation in small decision-making groups. In T. Parsons, R. F. Bales, et al. (Eds.), *Family, socialization, and interaction process*. Glencoe, Ill.: Free Press, 1955.

BROWN, J. F. *Psychology and the social order*. New York: McGraw-Hill, 1936.

CARTER, L. F. Recording and evaluating the performance of individuals as members of small groups. *Personnel Psychol.*, 1954, **7**, 477–484.

CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.

HOLLANDER, E. P. Conformity, status, and idiosyncrasy credit. *Psychol. Rev.*, 1958, **65**, 117–127.

HOLLANDER, E. P. Some points of reinterpretation regarding social conformity. *Soc. Rev.*, 1959, **7**, 159–168.

HOMANS, G. C. *The human group*. New York: Harcourt, Brace, 1950.

MOORE, O. K., & BERKOWITZ, M. I. Problem solving and social interaction. *ONR tech. Rep.*, 1956, No. 1. (Contract Nonr-609(16), Yale University Department of Sociology)

(Received August 17, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 370–379

CREATIVITY AND PRIMARY PROCESS:

A STUDY OF ADAPTIVE REGRESSION[1]

FRED PINE AND ROBERT R. HOLT
Research Center for Mental Health, New York University

BEGINNING with Freud (1921, 1953), but especially since Kris' explorations of creative processes (Kris, 1952; also Pappenheim & Kris, 1946), psychoanalytic theorists have described relationships among creativity, primary process (Freud, 1955), and a particular kind of ego control that permits adaptive use of primary process. The concept "regression in the service of the ego" (Kris, 1952; see also Schafer, 1958) refers to a momentary and at least partially controlled use of primitive, nonlogical, and drive dominated modes of thinking in the early stages of the creative process. The present study is one of the degree to which characteristic individual modes of expression and control of primary process in the Rorschach test are related to the quality of productions created in a variety of experimental tasks.

The study of creativity can begin from at least three different standpoints: the process, the product, and the person. Kris' work centers primarily though by no means entirely on the creative *process*, on the dynamic events surrounding the creative act itself. At the moment of creation, thinking is more fluid and often more drive dominated and less logical than usual; it is generally pervaded by intense emotionality—feelings of omnipotence, intense pleasure and/or pain (cf. Milner, 1957). Thinking closer to the primary process type is thus involved, but in the service of a productive act geared eventually to the creation of a communicable and socially valued product. The oft-described phases of creativity (cf. Kris, 1952; Rapaport, 1951) reflect this total process: the more primitive thinking characterizes the "inspirational" phase while the more controlled thinking characterizes the "elaborative" phase.

In a recent research, Pine (1959) studied the *products* of creative activity, employing the concepts of drive expression and drive control. The quality of productions—of essentially uncreative subjects (*S*s)—in a literary task (TAT) and in a task of scientific hypothesis formation varied with the amount and control of drive content in the product. Not only did the better productions include more drive content, but in addition such content was better controlled than in the poorer productions. Controls were reflected in the effective integration of the drive content into the total created production as well as in modulation of drive intensity and socialization of drive aim.

By contrast, the present study approaches the problem of creativity—and uncreativity—from the standpoint of the *person*, the creator. It involves a number of assumptions, from which three hypotheses were generated. We assume, first, that it is possible to score quantitatively evidences of the primary process that are apparent in a finished thought product—in the present case, in Rorschach test responses. Further, we assume that people differ in the degree to which they permit their responses to the Rorschach test to manifest recognizable signs of the primary process, or are unable to prevent its emergence in their responses (processes which we call, for short, the expression of primary process) and in the effectiveness of their control over such expressed primary process content. Finally, we assume that a score taking into account both the amount of expression of primary process and the effectiveness with which it is controlled constitutes an operational measure of adaptive regression (Kris' regression in the service of the ego). What we cannot assume, however, and what this study undertakes to explore, is the extent to which styles of expression and control of primary process on the Rorschach test are generalized and apply to a variety of tasks not all of which seem to require primary process modes of thinking.

Hypothesis 1. *S*'s tendency to express more or less primary process is unrelated to the quality of the productions he creates in various tests of imagination.

---

[1] This research was supported by a grant from the Ford Foundation.

Hypothesis 2. *S*'s tendency towards effective control over expressed primary process material is positively related to the quality of his created productions, effective control being associated with higher quality.

Hypothesis 3. *S*'s tendency towards adaptive or maladaptive regression to primary process modes of thinking (based on an evaluation of the combined expression and control tendencies) is positively related to the quality of his created productions, adaptive regression being associated with higher quality.

These hypotheses stem from an assessment of the relative significance of control—in contrast to expression—of primary process for creativity. Amount of expression per se takes no account of the control processes required for creative work. It is expected that *S*s who produce a good deal of primary process material on the Rorschach test with consistently good control will give the highest quality productions in the tests of imagination (Hypothesis 3). It is expected, further, that individual differences in control over primary process expression will be related to the quality measures, even apart from consideration of the amount of expression (Hypothesis 2). This is because almost all *S*s express at least some primary process content on the Rorschach test, because the tests of imagination that were used by and large do not require a capacity for expressive thinking comparable to that required for the production of, say, a major work of literature, and because we did not expect our sample of college students to contain many persons who would regress pathologically.

In testing these hypotheses, the study has one additional aim: to provide validation for Holt's system for scoring expression and control of primary process material in the Rorschach test (Holt, 1959; Holt & Havel, 1960). Previously, this system has been shown to provide an effective predictor of response to perceptual isolation, particularly in terms of the operation of primary process thinking during isolation procedures (Goldberger, 1958; Goldberger & Holt, 1961).

METHOD

Subjects. *S*s were 13 male and 14 female undergraduates at a large urban university. They were a group of volunteers who were intensively studied through clinical and experimental research techniques by numerous investigators for about one year. *S*s had been selected for the research program on the basis of (*a*) approximate indications on the MMPI of emotional stability, and (*b*) intelligence (upper 60% of entering college freshmen) on the Ohio State Psychological Examination. They have no particular excellence in creative arts or science nor, with two exceptions, do they have any professional commitment to these areas. *S*s are the same as those used in Pine's (1959) related study. The present study focuses on seven tests administered to all *S*s. "Creativity" scores were derived from each test, but the nature of this score varies with the particular test; in aggregate they are referred to as the "quality" scores.

Thematic Apperception Test (*TAT*). The TAT was used in investigating the quality of literary productions. The test was administered individually. Standard instructions were given and *S*s were told to take about 5 minutes per story and, on Card 11, to "tell a fairy tale such as you might tell a child." Nine stories per *S* were considered here, those given to Cards 1, 4, 11, 13MF, 16, 2, 18GF, 19, and 18BM (males) or 15 (females). Stories were recorded verbatim, the first five tape recorded and the last four written out by *S*.

The stories were rated on 12 attributes that were expected on a priori grounds to reflect the *literary quality* of the story: length, figures of speech, vocabulary, variety in sentence structure, naming characters, use of direct speech, characterization, time perspective, unity and continuity, genuineness, originality, and specification.[2] Fuller descriptions of these variables are given elsewhere (Pine, 1959); an attempt was made to control for length as much as possible in the latter 11 variables. Each of these story attributes was rated by one rater; five raters were used in all.[3] Although it was necessary to use each rater for more than one rating, only ratings which were unlikely to be confused with one another (e.g., variety in sentence structure and

---

[2] We wish to acknowledge our indebtedness to D. J. van Lennep, whose work in this area suggested some of these variables for analysis.

[3] One of the authors (FP) and four other raters carried out these ratings. Only FP knew anything about *S*s' scores on the other creativity tests, and he rated only the objective variables on the TAT (length, number of stories with direct speech, number of stories in which characters are given names). One of the TAT raters had some share in the ratings of the independent Rorschach variables. Although this rater had only a part share in rating the TAT (rating 2 of the 12 variables) and only a part share in rating some of the Rorschach protocols, correlations between the TAT quality scores and the Rorschach independent variables for handling of primary process were carried out with her ratings of the TAT excluded from the TAT score to check on any possibility of rating contamination. The correlations obtained did not differ in pattern or in statistical significance from those described below (Table 1) for the total TAT score.

We wish to express our acknowledgment to the many members of the staff of the Research Center for Mental Health who carried out various rating tasks.

characterization) were assigned to one rater. *S*s were ranked on each of the 12 story attributes and final ranks were derived by summing the 12 for each *S* and reranking the totals. The final combined ranking was used in all data analyses. The scoring reliability of the final literary quality scores was determined by the intraclass correlation technique (Ebel, 1951) applied to the 12 rankings; reliability of the mean score was .82 for the total sample.

The Science Test. The Science Test consisted of two problems, each posing a scientific dilemma which *S* was asked to resolve. One item (#2) was drive related and was generally felt by *S*s to be far the more fanciful while the other (#1) was more neutral. *S*s were required to write out three theories to account for each hypothetical situation. Instructions were to make the theories complete and realistic (i.e., probable) and to use imagination in a down-to-earth scientific manner. There was no fixed time limit, but no *S* took longer than one hour. The test called for skill in forming imaginative hypotheses; analysis and synthesis of information were only minimally involved and hypothesis testing was not involved. The problems were as follows:

1. A general idea of the way that coal forms is first that trees fall and decay; then owing to certain upheavals of the earth, land gets on top of them and over time the pressure of the earth turns the original material into coal. There are no trees in Antarctica, yet I want you to assume that coal has been found there. How could this be accounted for?
2. It is known that the differentiation of male and female sexes came about fairly early in animal evolution. That is, even in the lower animals male and female sex organs are in different individuals—there are separate male and female sexes. I want you to assume that recently a very advanced animal species, almost like the human species, has been found in which all individuals have both male and female sex organs. How could this be accounted for?

Ratings for *quality of theories* (on a four-point scale) were based on the specific theories given by *S* as well as on the general modes of scientific thinking (parsimony, a questioning attitude, use of known facts, etc.) reflected in the response. To test interrater agreement, Spearman rho between scores derived from each of the two raters' ratings was used; rho was .89 for males and .62 for females.[4] Findings from this two-item test can be regarded as suggestive only; the internal consistency of the test is nil (interitem rho on the quality scores is .05 for males and −.34 for females), and findings are presented for each item separately.

The Humor Test. The Humor Test consisted of 15 cartoons for which *S*s invented captions. It was administered individually with no time limit. *S*s were asked to write out responses and to make the captions as funny as possible. Captions were rated on an eight-point scale for *humorousness.* Five raters rated the captions given by female *S*s and these five plus two more raters rated those of male *S*s.[5] Each rater's 15 ratings for each *S* were summed to give a final score. Interrater agreement by the intraclass correlation technique (reliability of the mean rating) was .86 for the males and .87 for the females.

The Animal Drawing. Each *S* was asked to draw a "fantastic animal," making it as unusual as possible. Two judges independently ranked the drawings on *drawing ability* and on *originality.*[6] Interrater reliability (rho) for drawing ability was .90 for males and .84 for females; for originality, agreement was lower (rho .29 for males and .42 for females). Disagreements were discussed until a final ranking was achieved for each variable. The final ranks for the two variables are statistically independent (rho −.01 for males and .22 for females).

The Brick Uses Test. The Brick Uses Test (Guilford, Frick, Christensen, & Merrifield, 1957) required *S* to list as many uses as possible for an ordinary brick in 5 minutes; scores were based on the number of shifts in the type of use given. Guilford's scoring manual permits derivation of a score having high loading on his *spontaneous flexibility* factor. Interrater agreement in the present sample was 89%.[4]

The Consequences Test. The Consequences Test (Guilford et al., 1957) consisted of four items, each describing some hypothetical event (for example, the sudden destruction of all books). *S*'s task was to list possible consequences of the event (time limit: 5 minutes per item). Guilford's manual gives a score having high loading on his *originality* factor. Interrater agreement here was 86%.[4]

The Rorschach Test. The Rorschach Test was used to derive Rorschach creativity scores as well as rankings on the independent variables: expression and control of primary process.[7]

---

[4] Both Science Test items were rated by one of the authors (FP) and one other rater. These two raters also rated the Brick Uses and Consequences Tests (to be described below). Although the author in all cases knew which *S* had produced which response, the other rater never knew this and had no way of extending a halo effect across his several ratings. Ratings were carried out by the two raters working independently and final scores were obtained through discussion and reconciliation of differences. In no case did either rater have any knowledge of the ratings of the independent variables on the Rorschach.

[5] The seven raters (none of these the authors) were all unaware of which *S*s had produced which responses (and even of which responses were given by a single *S*). Although three of the seven had rated other creativity tests, the *S*s' names had been removed and there was no possibility of halo effect. One of the seven raters had a partial share in rating some of the Rorschachs, but no contamination was possible since he did not know which *S* he was rating on the Humor Test.

[6] The two raters (not the authors) were each ignorant of the Rorschach scores and of which *S* had produced which drawing. These raters each contributed to the ratings of the Humor Test but to none of the other tests.

[7] Ratings of these three Rorschach variables were carried out as follows: Each protocol was scored for all variables by two raters working independently and then coming to a consensus. The two raters varied from protocol to protocol, in each case selected from a total pool of seven raters. Any problems they had in arriving at a consensus were later discussed with one of the

The creativity rating: Each Rorschach response was scored on a four-point scale. The rating is based upon a composite of statistical infrequency, sensitive use of determinants, good form level, appropriate elaboration, and a quality of richness (see Holt, 1959). To avoid contamination with the primary process scores, the *creativity* ranking in the present sample was based on the average rating of only those responses that received no primary process scores.

Amount of expression of primary process: Holt (1956; Holt & Havel, 1960) has described a theoretical rationale and an extensive scoring manual for scoring primary process manifestations in the Rorschach. All responses that have aggressive derivatives, libidinal derivatives, anxious content, or various nonlogical or bizarre formal qualities are scored. The final ranking for amount of primary process is based upon the proportion of *S*'s responses that have scorable primary process elements, double weighting for extreme instances (Level 1 as compared to Level 2 in the scoring manual). Holt and Havel (1960) report interrater agreement well into the .90's (rho) for over-all scores on amount of expression of primary process.

Effectiveness of control over primary process: Holt and Havel (1960) describe procedures for rating the "defense-demand" and the "defense-effectiveness" of each response scored for primary process. Defense-demand, rated on a six-point scale, is defined in terms of the acceptability of the response in ordinary social communication; thus, for libidinal content, for example, the response "two people holding hands" requires almost no special controls while "two people having intercourse" does. Defense-effectiveness, also rated on a six-point scale, is defined in terms of the success of the controlling factors associated with the primary process material. Such things as the form level, the associated affective response of *S*, the intellectual or esthetic context of the response, and *S*'s evasiveness or confusion are considered. Thus, the composition, "This part looks like a horse but the rest makes me think it's a man" is a less effectively controlled formal primary process response than "a centaur," which refers successfully to the social reality of a myth, puts the response at a safe distance, and makes a condensation-like response sensible. Interrater reliability of the mean defense-demand and defense-effectiveness scores was computed for the male sample; rho's were .84 and .74, respectively.

Final scores for each *S* were derived from the defense-demand and defense-effectiveness scores by a procedure developed by Goldberger (1958). For each *S*, responses where defense-demand was high and defense-effectiveness was good were tabulated; similarly, responses where defense-effectiveness was poor, whatever the defense-demand, were tabulated; responses with good defense-effectiveness but low defense-demand were omitted. Final individual scores, which were then ranked, were based on the ratio of well-controlled responses (high demand, effective defense) to poorly-controlled responses (ineffective defense, whatever the demand).

Adaptive vs. maladaptive regression: This score is based on the combined amount of expression and effectiveness of control scores described above. It was developed and used with good predictive success by Goldberger (1958), and is described in detail by Holt and Havel (1960). In brief, it produces a ranking as follows: at one extreme are *S*s who are high on both amount of expression and effectiveness of defense (controlled or adaptive regression to primary process modes of thinking); at the other extreme are *S*s who are high on amount of expression but low on effectiveness of defense (uncontrolled or maladaptive regression to primary process modes of thinking); in the middle are *S*s who are low on amount of expression (constricted control).

Results

Before studying the hypotheses of the research, we examined some of the characteristics of the rankings of Rorschach primary process. Rank correlations between the measures of expression and control of primary process were computed. The rho's were negligible (.01 for the males and .24 for the females). These correlations indicate the statistical independence of the two variables as measured here; below, we treat them as independent in correlating each separately with the quality scores. The low correlations cannot, however, be taken to reflect functional independence in the individual. Rather, we would expect that expression and control of primary process would generally be integrated in some way in an individual's functioning—but in an idiosyncratic way. It is just these idiosyncratic possibilities in the relation between the expression and control of primary process that permit statistical independence in measurement.

The combined (adaptive regression) Rorschach score correlates far more strongly with

---

authors (RRH) in a meeting of all seven raters and a final decision reached. Although the Rorschach creativity scores were derived from only those responses where primary process was not scored, it is apparent that one cannot argue that these were uninfluenced by primary process scoring in the same protocol. The Rorschach creativity score turns out to be one of the poorest correlates of the independent Rorschach variables for handling of primary process, however, and makes no significant contribution to the results (see Table 1). The two primary process ratings (expression and control) were not independently rated although they turn out to be statistically independent (see below). Of the pool of seven raters, five (including RRH) did no other ratings for the study. One was one of the seven raters for the Humor Test (where he was "blind" to *S*) and one was one of the five TAT raters (and, as described in Footnote 3, omitting her ratings made no difference in the results).

the control score component (rho is .90 for males and .79 for females) than with the expression score component (rho is .11 for males and .13 for females). Thus, on methodological grounds alone, one would expect findings for the control and the combined scores to be more similar to each other than either is to findings with the score for amount of expression. We cannot, therefore, make independent tests of Hypotheses 2 and 3 but can only look to see the directions of change (if any) in the findings for the control score when the score for amount of expression is combined with it.

Hypotheses 1, 2, and 3 state that the quality of a person's created productions is unrelated to his tendency to express more or less primary process material on the Rorschach but is related to his tendency towards more or less effective control over expressed primary process content and to a combined score (adaptive vs. maladaptive regression) taking both amount of expression and effectiveness of control into account. Thus, it was expected that the quality scores on the various experimental tests would be correlated with the control and with the combined scores but not with the score for amount of expression alone. Results are presented in Table 1.

As indicated in Row 1 of Table 1, an over-all creativity score derived for each *S* correlated significantly with the control scores for both males and females and with the adaptive regression score for males; the correlations with the amount scores, though positive, are negligible. The over-all scores were obtained by summing the ranks for each *S* on eight of the nine individual test measures and reranking the totals; to avoid double weighting the animal drawing, and since the ratings of originality had poor interrater reliability on this test, only the ranking of drawing ability was used in computing the over-all scores. In view of the variations in the subtest correlations with the Rorschach scores for the female sample (to be described below), there is some question as to the representativeness of the over-all scores for the female sample. The validity of these scores was therefore investigated.

Staff members used qualitative data from tests and life histories to *Q* sort all *S*s on

TABLE 1

RANK CORRELATIONS BETWEEN TEST MEASURES OF CREATIVITY AND THE RORSCHACH SCORES FOR PRIMARY PROCESS

Creativity Score	Males (*N* = 13)			Females (*N* = 14)		
	Amount	Control	Adaptive Regression	Amount	Control	Adaptive Regression
Total Creativity: over-all score	.13	.80**	.90**	.21	.52*	.28
TAT: literary quality	.05	.74**	.83**	.59[a]	.27	.26
Science (neutral): quality of theory	−.17	.19	.08	−.32	.34	.24
Science (drive): quality of theory	.11	.43	.40	.64[a]	.11	.01
Rorschach: creativity	.17	.27	.31	−.22	.21	.22
Humor Test: humorousness	−.03	.13	.19	.32	.10	−.13
Animal Drawing: originality	−.14	.55*	.49*	.22	.38	−.05
Animal Drawing: drawing ability	−.04	.69**	.68**	.24	.39	.33
Brick Uses: flexibility	.24	.36	.48*	−.02	.44	.18
Consequences: originality	−.04	.59*	.74**	−.16	−.02	−.03

[a] Contrary to hypothesis. Significant at 5% level; two-tailed test.

* Significant at 5% level; one-tailed test.

** Significant at 1% level; one-tailed test.

various personality variables, which were then subjected to a syndrome analysis (Horn, 1944); syndrome scores were then converted to normative ranks (cf. Block, 1957). These rankings are partially contaminated since two of the eight component tests making up the over-all score (the TAT and the Rorschach) were used in evaluating *S*s for the *Q* sorts. With recognition of this qualification, three rankings were selected on a priori grounds to check the validity of the over-all scores. For the females, the over-all creativity score correlated with "creativity: needs for achievement, creative construction, love of novelty, sentience" (.63), with "clear and effective communication vs. blocked and uncommunicative" (.73), and with "vivid imagery vs. minimal and stereotyped fantasy and clings to reality" (.56); the corresponding correlations for the males were .61, .80, and .58, respectively. All these correlations are significant at the 5% level or better (two-tailed test). Thus, the over-all scores have some measure of validity. Be this as it may, valid test differences may still be obscured by these over-all scores.

Correlations between the Rorschach variables on the one hand and the quality scores on the individual tests on the other are presented in the remainder of Table 1. For the males, the correlations test by test parallel those with the over-all score. In each of the nine measures, the quality score correlates more highly with the Rorschach control score than with the amount score. In four instances, the correlation with control is significant and in all cases it is positive; in five instances, the correlation with amount is negative and in all cases it is negligible. The score for adaptive vs. maladaptive regression tends to give slightly higher correlations than the control score; the correlations are significant in five instances. There is little patterning among the tests that do and do not produce significant correlations; tests as different as the TAT, the animal drawing, and the Consequences Test give the highest correlations.

For the females, there is much more variation in the separate test correlations. Only two correlations are significant, and these—surprisingly—are both with the amount of primary process expression. Correlations with control and with the adaptive regression score are higher than those with amount in only about half of the cases.

Since other data available on this same female sample suggest that the adaptive regression score is not a meaningful one for this particular group (see Holt & Havel, 1960), and since this score produced no significant correlations in the present sample, we focus then on the two simpler (and independent) Rorschach scores, amount and control, for the females. The major difference in results between males and females is in the correlations with the amount score. For both groups, the correlations with control tend to be consistently positive (except for the Consequences Test for females) although the correlations for the females are smaller; that the females had less variation in their control scores may account for some of this difference. The wide range in correlations with amount in the female group, from strong positive to moderate negative, sharply differentiates the two groups.

The correlations nevertheless seem to fall into a pattern for the females. The tests which, at least in appearance, permit more open-ended and expressive verbal responses, and in particular permit more drive expression in the responses, are just those where the correlation for the female *S*s is higher with the Rorschach score for amount of primary process expression than with the control score. This category includes the TAT and the drive related Science Test item (and perhaps the Humor Test).[8] Why should it be that, for our college girls, individual differences in amount of primary process expression are at times better predictors of the quality of imaginative productions, and individual differences in control over expressed primary process material are at other times better predictors, while in the males the individual differences in control over primary process expression are consistently the better predictors of the quality scores? A critical point here, we suggest, is that people vary in their automatic (not necessarily thought out) approach to a task in terms of the implicit assumptions they hold regarding what the task and the experimenter requires of them. Thus, for tests taken in an experimental situation and in a university setting, *S*s may implicitly "direct" (or "set") themselves to give highly controlled imaginative productions; in this case, individual differences in the capacity to exert effective control over imaginative productions would get reflected in the quality of the finished product. In contrast, given such tests as the TAT, and in spite of the university surroundings, *S*s may at times approach test tasks as though the appropriate response were to give a fanciful and freely expressive response, even if the response would not be too well controlled; in this case, individual differences in the capacity for fanciful expression would be reflected in the quality of the finished product. In our data, it is as though the males consistently had the former (control) response set while the females had at times the former and at times the latter response set.

---

[8] On the face of it, the Rorschach test is also such an open-ended test and permits drive expression. It was not used in this analysis, however, because the Rorschach creativity score is based on only those responses where primary process is not scorable in the response and because a control factor (form level) is built into the rating criteria.

Why these college girls might have demonstrated such duality in their implicit approach to the tasks is of interest, and may reflect a greater field dependence (since they vary with the ostensible task "demand"), or a bimodal identity including both more expressive and more controlled aspects (feminine-masculine) with neither being consistently more dominant. Be this as it may, these speculations led to some expectations regarding the data. If the males responded to even the more fanciful tests in a way reflecting the effectiveness of their controls, and if the females responded to these tests in a way reflecting their potential for fanciful expression, then there should be two ways to get high (or low) quality scores on these tests: either by effectively (or ineffectively) controlled imaginative productions or by highly (or minimally) expressive productions. The males should tend to give the former, and the females the latter; but this should not be the case for the tests which "demand" more specific and controlled productions. Some indirectly relevant data were available for two of the more expressive tests (TAT and drive related Science) and for one of the more controlled ones (neutral Science). *S*s had been ranked for the *amount* of drive content in their TAT and Science productions and also for the degree of *control* over expressed drive content (see Pine, 1959, 1960); control was reflected principally in an integration of the drive content into the core of the created product. These ratings of expression and control of drive content are not identical with the ratings of expression and control of primary process on the Rorschach because of test differences and because the Rorschach ratings included formal as well as drive indices of primary process (although the latter predominated); however, what is more important, they do reflect expression and control within these created productions themselves and thus provide a better test of the hypothesis that there are two kinds of productions which could get high scores. It was expected that the quality scores on the TAT and the drive related Science item would correlate more highly with control of drive for the males than for the females and more highly with amount of drive expression for the females than for the males; this difference was not expected to appear in the neutral Science item. Correlations (rho) are presented in Table 2.

Correlations with amount of drive are higher for the females than for the males on the TAT and the drive related Science item but not on the neutral Science item. Correlations with the control scores are higher for the males than for the females on the TAT but the reverse is true on the neutral Science item. (Throughout, larger group differences appear in the Science items than in the TAT.) Thus, there is some indirect support for the proposition that the females achieved high (or low) quality scores on the more open-ended, expressive tests in a somewhat different way from

TABLE 2

RANK CORRELATIONS BETWEEN THE QUALITY SCORES AND EXPRESSION AND CONTROL OF DRIVE CONTENT IN THE TAT AND SCIENCE TEST

	TAT		Science (drive)		Science (neutral)	
	Amount	Control	Amount	Control	Amount	Control
Males	.60	.42	.09	.05	.12	.06
Females	.77	.28	.52	.04	−.36	.64

the males; individual differences in drive expression were more significant in the females than in the males on these tests.

The correlations shown in Table 2 are also of interest for their parallel to the correlations of Table 1. In both, for the male sample, the TAT produces a significant correlation with the control score (although the correlation with amount is higher in Table 2), while the two Science items produce no significant correlations. The variation in the correlations for the females found in Table 1 is repeated almost exactly in Table 2. In the TAT and the drive related Science item, all correlations are positive and the correlation of the quality score with amount is much higher than with control; in the neutral Science item, the correlation with control is positive and with amount is negative. Thus, with amount and control scores derived from the TAT and the Science productions themselves as well as with the parallel scores on the Rorschach test the same sex differences and patterning of correlations are obtained. These findings support the validity of these patterns of difference in the present sample.

DISCUSSION

The results are suggestive in several ways. First, the way in which primary process material is handled on the Rorschach test is related to the quality of imaginative production in a variety of test situations. That control over primary process expression, rather than gross amount of expression per se, is related to the quality of created products is indicated by all of the data of the male sample and by results with the over-all creativity score for the females. These results give general support to the hypotheses of the study, and they indicate further that Kris' concept of regression in the service of the ego can be subjected to empirical study—that predictions generated by this psychoanalytic concept can be studied with operationally defined scores and quantitative techniques.

The data for our male sample, where scores for adaptive regression and for controlled primary process expression could predict quality scores on tests of imagination even where primary process thinking was not evidently required by the tests, are of additional interest. The tests used here have certain parallels to the model of primary process thinking: thus, the TAT requires a "make-believe," nonliteral kind of thinking and also the use of drive content (for motivation of characters); the role of primary process in humor has been discussed by Freud (1938); even with the Brick Uses there is a parallel to primary process in that high scores are attained by *shifts* in the direction of thinking, by a style that contrasts with the directed thought of the secondary process. But these parallels between the test requirements and primary process thinking are at best tenuous in most or all of the tests used. We would argue, in contrast, that primary process thinking is not directly implicated in the production of responses to these tests, but rather that modes of expression and control of primary process (perhaps beginning in early childhood) become generalized as broad cognitive styles (cf. Klein, 1950) which are reflected in all areas of thinking. Thus, where tests require flexible and original thinking, and where the control style permits this, thinking can be flexible even apart from the (hypothesized) original relation of control directly to primary process thinking.

But there is an important caution to be drawn from our results. The subsample differences indicate that there are limitations to the extent to which we can generalize from these data. Kris' discussion of the relation of regression in the service of the ego to creativity can be viewed as a broad approximation—an approximation subject to important differences in samples that vary as to age, sex, the absolute level of creativity of the *S*s involved, and the nature of the creative production. The most we can say from our data is that expression and control of primary process are relevant to an understanding of the creative process, but that the specific nature of the relationships involved may vary and awaits much further research for detailed specification.

The sex differences in our data are of interest for their implications regarding the relationship between "creativity" on the present test measures and creativity as it is more naturally found in the productive arts and sciences.

First: Do the two significant correlations between quality scores and *amount* of primary process expression (rather than with effectiveness of control) indicate that some highly rated productions were quite expressive but seriously lacking in controls? This would contrast to an "ideal" of creative productivity as both personally expressive and sufficiently controlled to be communicable and potentially socially valued. Whether or not some truly creative productions can be really uncontrolled, or whether the nature of controls in some areas (e.g., painting, dance) is simply different from and perhaps more subtle than the intellectual verbal controls required by most of our tests, the essential point is that our data do not really suggest that some high quality productions here were uncontrolled. The *S*s who gave these results were "normal" college girls. It can be assumed that some control factors were included in *all* of their expressions of primary process, and the Rorschach scores support this assumption. Furthermore, these girls are obviously all sufficiently well controlled to function effectively in school and generally, and presumably they would be able to impose control when the adaptive requirements of a task require this (as in the tests of imagination) even if some of them could relinquish control somewhat on the Rorschach.

Second: The introduction of the idea that our male and female *S*s had implicit but differing response sets which led them to approach the various tasks differentially requires comment. Certainly this concept is relevant for our present sample where sudden "inspiration" had little or no part in the production (upon demand) of imaginative products and where varying expectations and goals were likely to be activated by the entire experimental setting. But does such variation in approach to productive work, variation in terms of more expressive or more controlled responsiveness, apply in the creative arts and sciences? The concept certainly does not seem applicable where creative experiences come in off-guard moments, during sleep, or when the sudden inspiration is experienced as though coming upon the person from outside without his being responsible for it (cf. Ghiselin, 1952). But there is a way in which individual preferences for more expressive or more controlled productivity are relevant for the broader realms of creativity: in the way in which people, consciously or not, gravitate towards areas conducive to differing kinds of productivity. Thus, as the girls in the present sample may have responded with more or less fanciful and idiosyncratic expression in tasks which permitted such expression, so there seem to be differences in western culture in the areas where males and females have made their greatest contributions. A Sappho existed long before a Madame Curie, and it is probably still true that there is a greater proportion of women in the arts than in the sciences. Thus, alternative areas open for productive work in a particular social milieu may provide room for differing individual styles of creativeness.

SUMMARY

The present paper reports a study of relationships between expression and control of primary process material on the one hand and quality of productions in a variety of tests of imagination on the other. Scores reflecting the quality of imaginative productions were derived from the following tests: TAT, Science Test, Humor Test, Rorschach, Animal Drawing, and Guilford's Brick Uses and Consequences Tests. Holt's manual was used in assessing expression and control of primary process in the Rorschach test. *S*s were 13 male and 14 female undergraduates.

Amount of expression of primary process and effectiveness of control over such expression were shown to be statistically independent. It was expected that the quality scores on the tests of imagination would be unrelated to amount of primary process expression and positively related to the effectiveness with which such expression is controlled. Results on all tests for the males, as well as over-all scores for both groups, supported the hypothesis. The correlations for the individual tests for the females, however, showed many unexpected variations.

An hypothesis regarding the results with the female sample was advanced and explored—based on sex differences in implicit orientation to the tasks which produced relative differences in the expressiveness and control of responses in males and females. Implications of the findings in relation to creativity as it is encountered in natural, nonexperimental, settings were discussed.

REFERENCES

BLOCK, J. A comparison of ipsative and normative ratings of personality. *J. abnorm. soc. Psychol.*, 1957, **54**, 50–54.

EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, **16**, 407–424.

FREUD, S. *Delusion and dream*. London: Allen & Unwin, 1921.

FREUD, S. Wit and its relation to the unconscious. (German ed., 1905) In A. A. Brill (Ed.), *Basic writings of Sigmund Freud*. New York: Modern Library, 1938.

FREUD, S. Formulations regarding the two principles in mental functioning. (German ed., 1911) In, *Collected Papers*. Vol. IV. London: Hogarth Press, 1953. Pp. 13–21.

FREUD, S. *The interpretation of dreams*. (German ed., 1900) New York: Basic Books, 1955.

GHISELIN, B. *The creative process*. Los Angeles: Univer. California Press, 1952.

GOLDBERGER, L. Individual differences in the effects of perceptual isolation as related to Rorschach manifestations of the primary process. Unpublished doctoral dissertaion, New York University, 1958.

GOLDBERGER, L., & HOLT, R. R. Experimental interference with reality contact: Individual differences. In P. Solomon (Ed.), *Sensory deprivation*. Cambridge: Harvard Univer. Press, 1961, in press.

GUILFORD, J. P., FRICK, J. W., CHRISTENSEN, P. R., & MERRIFIELD, P. R. A factor-analytic study of flexibility in thinking. *U. Sth. Calif. Psychol. Lab. Rep.*, 1957, No. 18.

HOLT, R. R. Gauging primary and secondary processes in Rorschach responses. *J. proj. Tech.*, 1956, **20**, 14–25.

HOLT, R. R. Manual for the scoring of primary process manifestations in Rorschach responses. (Draft 7) New York: Research Center for Mental Health, 1959. (Dittoed)

HOLT, R. R., & HAVEL, JOAN. A method for assessing primary and secondary process in the Rorschach. In Maria A. Rickers-Ovsiankina (Ed.), *Rorschach psychology*. New York: Wiley, 1960. Pp. 263–315.

HORN, D. A study of personality syndromes. *Character & Pers.*, 1944, **12**, 257–272.

KLEIN, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality*. New York: Ronald, 1950. Pp. 328–355.

KRIS, E. *Psychoanalytic explorations in art*. New York: International Univer. Press, 1952.

MILNER, MARION. *On not being able to paint*. New York: International Univer. Press, 1957.

PAPPENHEIM, ELSE, & KRIS, E. The function of drawings and the meaning of the "creative spell" in a schizophrenic artist. *Psychoanal. Quart.*, 1946, **15**, 6–31.

PINE, F. Thematic drive content and creativity. *J. Pers.*, 1959, **27**, 136–151.

PINE, F. A manual for rating drive content in the Thematic Apperception Test. *J. proj. Tech.*, 1960, **24**, 32–45.

RAPAPORT, D. Toward a theory of thinking. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951. Pp. 689–730.

SCHAFER, R. Regression in the service of the ego: The relevance of a psychoanalytic concept for personality assessment. In G. Lindzey (Ed.), *Assessment of human motives*. New York: Rinehart, 1958. Pp. 119–148.

(Received August 17, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 380–388

EFFECTS OF STRESS ON GOAL SETTING BEHAVIOR

KUNAPULI U. RAO
Government Training College, Adnhra Pradesh, India

AND

ROGER W. RUSSELL[1]
Indiana University

IN an earlier paper one of us (Russell, 1953) discussed analogies between the concept of "stress" as it has been applied to the dynamics of physical, biological, and behavioral systems. Many of these are systems that tend to maintain relatively steady states over extended periods of time. External forces, stresses, impinging upon the system are balanced by changes, strains, within the system. More recently Davis (1958) has described this dynamic process in terms of "negative feedback": "[The] basic requirement [of negative feedback] is that some energy is taken from a later part of the system and introduced into an earlier part in such a way as to oppose the change produced there by the original input" (p. 10). The relation between input and output, or between stress and strain, holds within limits; beyond the limits, however, stresses may produce temporary or permanent distortions of the system.

Applying this model to behavior systems suggests that disorders of behavior may occur when responses are inadequate to compensate for changes in the systems produced by stresses impinging upon them. On a priori grounds it might be predicted that this state of affairs would exist when the organism does not possess the capabilities or skills required for an adequate response or when competition arises between two or more incompatible response tendencies, resulting in delay or failure in the occurrence of an adequate response. In the psychological literature the former condition has frequently been referred to by the generic term, "frustration," and the latter by "conflict." Experimental studies have shown that exposure of a variety of living organisms to either of the conditions may indeed produce behavior disorders (Russell, 1950); however, distortions of behavior systems are by no means inevitable results in such experiments. Distortion would not be expected to occur if feedback from the performance or output of a system altered the system in such a direction that its response capabilities were adequate to compensate for changes produced by future inputs of a similar nature. The present experiment was designed to study a familiar and frequently occurring situation in which behavior systems are altered in this way.

It is a commonly observed characteristic of human behavior that a person when frustrated by failure to achieve a goal he has set for himself tends to alter the goal in a direction more consistent with his actual performance. This is a process in which feedback, knowledge of results, from the output, performance, produces changes in the behavior system, i.e., changes in goals; the latter, in turn, affect the future response of the system. We have seen this goal setting process in action in experimental situations designed to impose stress upon our subjects; its effect has been to reduce or to eliminate predicted disorganization of behavior. In the light of clinical as well as experimental evidence it appears that goal setting is an important process in preventing disorganization of behavior and, therefore, deserves special attention in experimental psychopathology.

The process of goal setting may be studied experimentally using the familiar level of aspiration technique (see Frank, 1935b; Lewin, Dembo, Festinger, & Sears, 1944; Rotter, 1954). The technique has three major features: at the beginning of each trial the subject states the goal he expects to achieve during the trial; he then performs the required task; and, finally, he obtains knowledge of the results of his performance. Knowledge of results provides a feedback the effects of which on stated levels of aspiration and performance may be observed in subsequent trials. In the present experiment we have used this behavior system to study the effects on goal setting behavior of stress imposed on the system by failure to achieve stated goals. By controlling performance without the subject's knowledge it has been possible to vary systematically the number of successive trials during which failure occurred. We have also studied the aftereffects of the stress on levels of aspiration and on performance during a recovery period when failure was not induced experimentally.

---

[1] The research was conducted at the Department of Psychology, University College, London, England. We wish to express our appreciation to A. R. Jonckheere and A. Summerfield for their critical examination of our methods for analyzing results of the study and to the Headmaster of the school whose very kind cooperation facilitated the collection of our data.

The model in terms of which we have viewed the goal setting process permits certain predictions that should be testable under the conditions of the present experiment:

1. The goal setting process should lead to stated levels of aspiration that are consistent with performance and remain relatively stable as long as performance is stable. This is, of course, the familiar result of many studies employing the level of aspiration technique.

2. Stressing the behavior system by introducing repeated failure to achieve stated goals should upset the stability of the prestress levels of aspiration and lead toward the establishment of new levels. This prediction is also supported by evidence from a number of previous experiments in which it has been a consistent finding that level of aspiration tends to be raised after successful performance and lowered after failure.

3. There should be an inverse relation between duration of exposure to stress and level of aspiration such that the level decreases as the number of stress trials increases. We are particularly interested in the nature of this relation and our research was designed to provide an empirical description of it.

4. Return to prestress experimental conditions should be reflected in shifts of stated levels of aspiration in the direction of their original levels of stability. Here we are particularly interested in two characteristics of the recovery process: (*a*) how is rate of recovery related to the duration of previous exposure to stress; and (*b*) is there a tendency toward "overshooting" and oscillation such as may occur in other feedback systems (Davis, 1958; Selye, 1946)?

METHOD

Experimental Design

The general plan of the research was to obtain measures of level of aspiration and of performance before and after exposure to varying numbers of trials in which subjects consistently failed to achieve their stated goals. The plan also included measurement of level of aspiration during the stress trials; performance could not be measured during this phase of the research since the experimenter manipulated performance scores artificially in order to produce consistent failure to achieve levels of aspiration.

The experimental design included four phases, always presented in the same temporal order. It required four groups of experimental and four of control subjects, all groups matched initially on variables likely to affect behavior in the research situation. In Phase 3, the four experimental groups were subjected to "stress trials" totaling 10, 20, 30, and 40 trials, respectively, while the control groups performed corresponding numbers of "standard trials." In all other phases experimental and control groups received identical treatments.

Subjects

The subjects were 160 adolescent boys, chosen from a secondary technical school. The experiment was planned in two replications, the first consisting of 96 and the second of 64 subjects. The former were pupils of the third form during the school year 1952–53; the latter, pupils of the same form admitted during 1953–54.

Previous research (see Lewin et al., 1944; Rotter, 1942a) has shown that age, sex, socioeconomic background, past success and failure, and intelligence are major variables affecting goal setting behavior in the level of aspiration situation. Therefore, it was necessary to match the eight groups required by the experimental design on these variables. Possible effects of sex differences were controlled by using only male subjects. All boys selected as subjects belonged to families of the socioeconomic class, "manual workers" (Mosier & Hall, 1954). To control for possible effects of past academic success or failure only boys with satisfactory academic progress during the preceding school year participated in the experiment. Table 1 shows that there were no significant differences among the groups in their chronological ages, intelligence quotients,[2] nor in their initial scores on measures of the behavior which constituted the dependent variables in the experiment.

Apparatus and Instructions

The apparatus consisted of a slightly modified form of the Rotter Aspiration Board (Rotter, 1942b). It was used to present a simple motor task in which the subject was required to use a small cue to hit a steel ball along a groove cut into a wooden base block. Runners attached to both sides of the groove kept the ball in the channel. Marked along the two sides of the channel were 19 units, numbered in increasing order from 1 to 10 and then decreasing to 1. In the channel at points opposite the center of each unit was a circular indentation, in one of which the ball came to rest after each hit.

---

[2] Intelligence quotients were derived from scores on the Otis Self-Administering Test, Intermediate Examination, Form A, as adapted by the Australian Council for Educational Research.

TABLE 1

MATCHING OF CONTROL AND EXPERIMENTAL GROUPS

Variable	F^a	df	Mean
Chronological age (months)	0.57	7 and 152	166.6
Intelligence quotient	1.58	7 and 152	106.6
Initial level of aspiration	1.59	7 and 152	30.3
Initial performance	0.64	7 and 152	27.7

[a] None of these F ratios is significant.

In Phase 3 of the experiment it was necessary for the experimenter to control the point where the ball came to rest. This was accomplished by altering the relative positions of the runners along the sides of the channel, thereby narrowing the channel and jamming the ball. The apparatus was so designed that this could be done without the knowledge of the subject. One runner consisted of two pieces, one piece pivoted at its central end and the other end joined to a Bowden cable by means of a bracket. The second runner was also pivoted at one end and joined to the cable at the other. By turning the cable control handle to different extents the ball could be stopped at various points along the channel.

In accordance with the practice adopted by previous investigators using this task, five hits were taken to constitute one trial. The sum of the scores for the five hits provides a score that is more stable than scores for individual hits. Before each hit the steel ball was placed at a standard point at the subject's end of the channel. Each trial of five hits began with the subject's stated level of aspiration, i.e., the number of points he expected to obtain during the trial. The subject then hit the ball with the cue. After each hit, the subject and the experimenter could determine the score in terms of the number of the unit on the channel in which the ball came to rest.

Rotter (1942a, 1942b) has shown that the instructions given subjects regarding the task have important influences upon the results obtained. Therefore, the standardized instructions used in the present study conformed to those suggested by his research. Precautions were taken to be certain that the instructions were clearly understood. After they were given, the subject was provided an opportunity to practice. This practice consisted of 10 hits for each; it allowed the experimenter to observe whether or not the subject could apply the instructions correctly. Immediately after this practice period the main features of the instructions were repeated and a standard series of questions asked, the answers to which indicated whether or not the subject had understood. Further explanation was given if necessary.

Procedure

As far as possible subjects from the various groups were taken in a random order, but sometimes this order could not be adhered to because of absences, school examinations, etc. The experiment was conducted at the school in a private room made available for the purpose.

Phase 1: Practice. When a subject arrived, he was seated in a chair at the subject's end of the apparatus and given the general instructions, the 10 practice hits, and a review of the instructions as described above.

Three main steps were taken to insure that the subjects were as highly motivated as possible during the experimental period.

1. The experiment was introduced as a new game.
2. Subjects were told that their scores were measures of their general ability and would be seen by the Headmaster of the school.
3. Monetary prizes were offered and given when the study was completed, the amounts being set after consultation with the Headmaster and taking into consideration the extent of funds normally available to schoolboys as "spending money." Replies to a standard series of questions, put to each subject when his experimental session had been completed, suggested that all worked for good scores. They appeared to be more concerned with performing well in order to gain approval of the experimenter and the Headmaster than to obtain the monetary rewards.

Phase 2: Prestress. During Phase 2, the prestress phase, all subjects were given 15 standard trials of five hits each. The procedure for each trial contained three basic steps. The subject first stated his level of aspiration, the score he expected to obtain, for that trial; he then performed the task; and finally he was given knowledge of his performance scores.

Phase 3: Stress. Two independent variables, presence or absence of stress and number of stress trials, were introduced in this phase of the experiment. As stated earlier, stress was defined as trials during which the subject consistently failed to achieve his stated levels of aspiration. Performances during stress trials could be controlled since the experimenter manipulated the apparatus without the subject's knowledge, thus determining at which unit in the channel the ball would come to rest after each hit. All experimental groups were exposed to such stress, successive groups receiving 10, 20, 30, and 40 stress trials, respectively. The various control groups continued the standard procedure, as in Phase 2, for numbers of trials corresponding to their paired experimental groups.

Phase 4: Recovery. During the final phase of the experiment all groups were given 40 standard trials, following the procedure used in Phase 2.

Before leaving the experimental room each subject was asked not to discuss the experiment with anyone in the school and, as a token of his willingness to conform to this request, signed his name on the experimenter's score sheet. Examination of performance scores and of replies to questions at the end of the experimental session strongly suggest that the subjects kept strictly to their promises and that no information about the experiment leaked out. The possibility of losing one of the larger monetary rewards to another boy with prior information may also have helped in achieving this end.

RESULTS

Results obtained in the various phases of the experiment are described in the following paragraphs. Since our principal interest in the study was in the goal setting process, the data

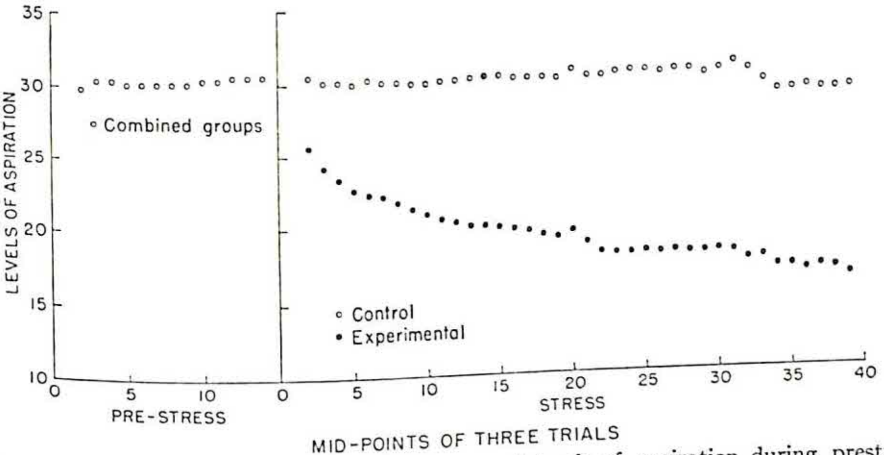

FIG. 1. Curves representing three-point moving averages of levels of aspiration during prestress and stress trials.

for level of aspiration and performance are presented separately. Throughout the report, "significance" is defined conventionally in terms of the .05 level of confidence.

Prestress

Goal setting behavior. A general picture of goal setting behavior when the subjects worked freely at the experimental task is shown by the curve for combined groups on the left side of Figure 1. This, as all other curves presented in the report, is expressed in terms of three-point moving averages in order to show clearly the trends of the data. Prior to the prestress phase of the study, all subjects received 10 practice hits; this practice was sufficient to have resulted in a level of aspiration that remained extremely stable for all experimental and control groups during the 15 prestress trials. Table 2 presents the mean prestress levels of aspiration for all groups; none of the very small intergroup differences was significant.

Performance. Figure 2 shows that performance also reached a stable state very early in the prestress trials. There were no significant intergroup differences in the mean performance scores, which are presented in Table 3.

Stress

Goal setting behavior. Levels of aspiration for all control groups continued to remain stable throughout the next phase of the study, when experimental groups were being subjected to stress. Each of the four control groups received different numbers of trials, i.e., 10, 20, 30, and 40, during this phase, but, in order to show a composite picture of their goal setting behavior, the scores have been combined in constructing the curve shown in Figure 1. There were no changes in the conditions under which the control subjects operated and, as Figure 1 indicates, their levels of aspiration continued at the same steady state level as during the prestress phase.

Conditions were changed for the four experimental groups, who received 10, 20, 30, and 40 successive trials, respectively, during all of which the subjects failed to achieve their stated levels of aspiration. As Figure 1 shows, the change was reflected immediately in the subjects' goal setting behavior. Stated levels of aspiration always decreased on successive trials, although at an apparently decelerating rate as the number of stress trials increased.

There are several possible ways of expressing these descriptive trends in numerical terms. The method chosen for the present analysis provided a "net score" calculated as follows: The mean level of aspiration for each subject in the prestress phase was calculated. Differences between this standard and levels at subsequent stages in the experimental session were determined as measures of the increase or decrease in level associated with the experimental treatment. The differences between such measures for an experimental subject and his corresponding control provided "net scores." Table 4 presents the median net decreases in levels of aspiration of the four experimental groups during the stress phase of the study. The fact that these derived scores for all groups were in the same direction, i.e., decreases, indicates that exposure to stress was associated in all groups with consistent alterations in goal setting behavior. The data in Table 4 also suggest that the magnitude of these decreases in aspiration level were related to the number of stress trials to which the groups were exposed. Jonckheere's (1954) distribution-free *k*-sample test against ordered alternatives was applied to the data, and this trend was found to be significant ($p < .01$).

The curve in Figure 1 suggests that the rate of decrease in level of aspiration tended to decelerate with increasing duration of exposure to stress. This curvilinear relation is shown in Table 5, which presents the mean net shifts in levels during successive blocks of 10 trials. The means were derived in the following way: For Block 1 the shift for each subject in the four experimental groups was obtained by finding the difference between his mean level during prestress trials and that during the first 10 stress trials. Similar calculations for corresponding control subjects provided measures representing their shifts. The differences between such measures for experimental and control subjects provided "net shifts." Inspection of the columns in Table 5 shows that, with one exception, the magnitude of these net shifts decreases consistently as the number of stress trials increases. Using analysis of variance for Blocks 1 and 2, and a *t* test for Block 3, it was found that there were no significant differences among the experimental groups in any one block. Therefore, the shifts in each block for all groups were pooled and an overall mean determined as shown in the final column of the table. Applying Jonckheere's (1954) test, the trend for the overall mean shift to decrease as the number of stress trials increased was found to be significant ($p < .01$). In other words, the effects of continued exposure to stress diminished significantly as the total number of stress trials increased.

It is, of course, impossible to present a similar analysis of performance scores during the stress phase of the study since the experimental design required that all trials for the experimental groups be influenced artificially by the experimenter in order to produce consistent failure. Without meaningful performance scores it is also impossible to present the derived "difference" or "goal discrepancy" scores (Frank, 1935a; Lewin et al., 1944) frequently reported by investigators to express the relation between a level of aspiration and the level of performance that preceded it. Since we are primarily interested in the goal setting process itself, such derived scores are not essential to our report.

TABLE 2

MEAN LEVELS OF ASPIRATION: PRESTRESS PHASE

Replication	Groups							
	C_1	E_1	C_2	E_2	C_3	E_3	C_4	E_4
1	30.1	30.8	29.4	30.6	31.2	31.6	29.8	31.0
2	30.7	31.2	29.4	29.6	30.8	29.3	28.3	29.8
Combined[a]	30.3	30.9	29.4	30.2	31.1	30.6	29.2	30.5

[a] The differences among groups are not significant ($F = 1.59$, $df = 7$ and 152).

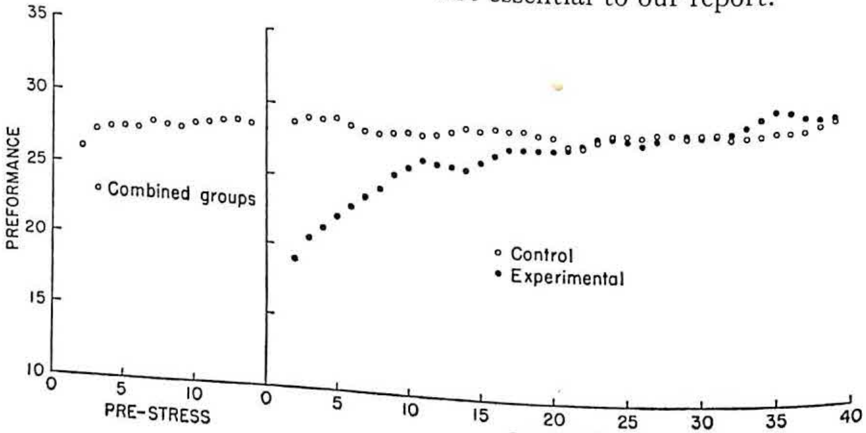

FIG. 2. Curves representing three-point moving averages of performance during prestress and recovery trials.

TABLE 3

MEAN PERFORMANCE SCORES: PRESTRESS PHASE

Replication	Groups							
	C_1	E_1	C_2	E_2	C_3	E_3	C_4	E_4
1	27.7	26.8	27.8	27.4	27.4	28.2	27.4	28.0
2	27.6	28.3	27.8	26.9	28.6	28.0	27.1	27.8
Combined[a]	27.6	27.4	27.8	27.2	27.9	28.1	27.3	27.9

[a] The differences among groups are not significant ($F = 0.64$, $df = 7$ and 152).

Recovery

Goal setting behavior. At the start of the fourth phase of the study changes in operating conditions were introduced for the experimental groups by allowing the subjects once again to work freely at the experimental task; control subjects still operated under the standard conditions that had prevailed for them throughout the experiment. Figure 3 shows that the levels of aspiration of the control subjects continued at the same steady state level as during the preceding two phases of the study; for the four control groups the steady state had persisted for 65, 75, 85, and 95 trials, respectively.

On the other hand, removal of the stress conditions introduced a rapid change in the stated levels of aspiration for all the experimental groups. The ascending curve in Figure 3 shows the general trends of the goal setting process. The process resulted in recovery to prestress levels of aspiration, levels which had characterized the subjects' behavior under conditions identical with those now existing and which corresponded to those maintained by the control groups throughout the experiment. The sensitivity of the goal setting process to the change in operating conditions is shown by the rapid shift from a three-point moving average of 16.1 at the end of the stress trials to an average of 20.8 at the beginning of the recovery trials.

Figure 3 is a composite curve constructed by combining scores for subjects in all four experimental groups. Although it shows general trends, it does disguise one kind of information that is clearly apparent when curves are plotted for each experimental group individually, i.e., the duration of the recovery process appears to be related to the number of stress trials imposed during the preceding phase of the experiment. This relation can be expressed numerically in several ways, all of which give similar results. The data we report here were obtained by a simple graphic method using curves plotted for each experimental group individually: the number of the trial on the abscissa corresponding to the first point on the recovery curve that equaled the prestress mean level of aspiration for the group was taken as representing the number of trials for recovery. The data are presented in Table 6. The trend toward increasing number of trials to recover with increasing number of stress trials which is apparent in these data was found to be significant when tested by Jonckheere's (1954) method and by analysis of variance: by both methods $p < .01$. A more detailed examination of the trend suggests that it is not rectilinear, but tends to approach an asymptote between 30 and 40 stress trials. This possibility was examined by t tests between all the various pairs of experimental groups, using data for the combined replications. The results of the tests are given in Table 7 and show that, as inspection of the basic data had suggested, all differences in number of trials to recover were significant except the difference between Groups E_3 and E_4, the groups with 30 and 40 stress trials, respectively.

We wish to call attention to one other feature of the recovery curve for the experimental groups. Examination of the curve in Figure 3 suggests that there was some oscillation and "overshooting" in the recovery process after the prestress levels of aspiration had been attained. Had we had some earlier indication of when in the recovery trials these phenomena might have occurred, we would have extended the recovery phase with the intention of determining if and when the phenomena disappeared and a final steady state reached.

TABLE 4

"NET DECREASES" IN LEVELS OF ASPIRATION OF EXPERIMENTAL GROUPS: STRESS PHASE

Groups	Number of Stress Trials	Median Score
E_1	10	7.5
E_2	20	9.2
E_3	30	10.2
E_4	40	11.7

TABLE 5

MEAN "NET SHIFTS" IN LEVELS OF ASPIRATION OF EXPERIMENTAL GROUPS DURING SUCCESSIVE BLOCKS OF TEN TRIALS: STRESS PHASE

Block	Stress Trials	Groups				Overall Mean
		E_1	E_2	E_3	E_4	
1	1–10	8.0	8.2	7.3	6.9	7.6
2	11–20	—	4.8	3.2	3.2	3.7
3	21–30	—	—	1.7	3.6	2.6
4	31–40	—	—	—	0.6	0.6

Performance. A general picture of changes in performance scores during the recovery phase of the experiment is presented in Figure 2. It seems clear that levels of performance of the experimental groups, as well as their levels of aspiration, had been affected by the changed conditions which operated for them during the stress trials. On the other hand, the performance of the control groups appear to have continued at the same steady state level as during the prestress phase of the study. Recovery from the effects of the stress trials occurred in all groups, the general nature of the recovery process having features in common with that of goal setting behavior under identical conditions.

Data for the experimental groups on the number of trials for recovery of prestress levels of performance have already been presented in Table 6. When analyzed in the same manner as were the level of aspiration scores, the performance data show a significant trend toward increasing number of trials to recover with increasing number of preceding stress trials. The trend tends to approach an asymptote between 30 and 40 stress trials. Examination of the recovery curve in Figure 2 also suggests that oscillation and "overshooting" occurred in the recovery process after the prestress levels of performance had been attained.

DISCUSSION

Before summarizing our results and conclusions we wish to discuss briefly three aspects of the present experiment.

Stress

Stress, as defined operationally in the present study, was induced by controlling the performance of our experimental subjects, without their knowledge, so that they failed to achieve their stated levels of aspiration during a predetermined number of successive trials. Basically this places the subject in an insoluble problem situation and exposes him to the "stress" that characterizes such situations (Russell, 1953). Berkeley (1952) has demonstrated that exposure to such stress not only affects behavior but also produces physiological changes associated with a wide variety of other forms of stress. One of these physiological changes involves the pituitary-adrenal cortical system (Selye, 1946); exposure to stress results in increased adrenal cortical activity, the increase being proportional to the degree of stress imposed (Hoagland, 1947; Pincus & Hoagland, 1943). In Berkeley's experiment subjects were required to perform the same task used in the present study, failure being controlled by preventing a subject, without his knowledge, from attaining a "minimum score." Effects of exposure to such stress were studied in terms of differences between rates of excretion of urinary 17-ketosteroids under prestress and stress conditions. The results showed that the extent of the discrepancy between aspiration and performance was directly related to the rate of 17-ketosteroid excretion, a multiple correlation coefficient of .70 expressing the magnitude of this relation. In our present experiment we have studied goal setting behavior under basically the same conditions as Berkeley; his findings aid in relating our results to the general concept of stress as it has been widely applied not only by psychologists, but also by research workers in other disciplines.

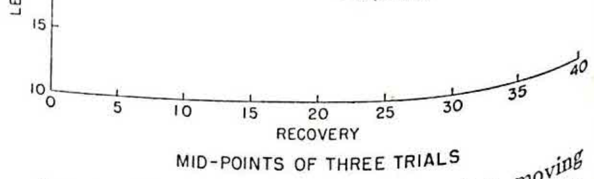

FIG. 3. Curves representing three-point moving averages of levels of aspiration during recovery trials.

TABLE 6

NUMBERS OF TRIALS FOR RECOVERY: LEVELS OF ASPIRATION AND OF PERFORMANCE: COMBINED REPLICATIONS

Group	Level of Aspiration	Performance
E_1	10.1	8.4
E_2	16.7	15.2
E_3	21.4	22.0
E_4	24.9	22.2

Lower Limits to Changes in Level of Aspiration under Stress

During the stress trials the trend in stated levels of aspiration was to decrease consistently as the number of stress trials increased. The decrease was at a decelerating rate, as if the goal setting process were approaching a new steady state level despite continued exposure to stress. Results of other studies (see Lewin et al., 1944) have suggested that there may be absolute limits for any situation, within which the goal setting process may function but beyond which levels of aspiration will not be set. Such limits would presumably be dependent upon other variables affecting the behavior system, e.g., age and sex of the subjects, their socioeconomic status, past success and failure, intelligence, etc. Our experiment was designed to control rather than to study these variables; therefore, we can provide no information on how they may have determined a lower limit for levels of aspiration during the stress trials. It is a reasonable hypothesis that the asymptote which the levels appeared to be approaching may have been set by such a lower limit.

"Overshooting" and Oscillation

There is a tendency for feedback systems to respond to change by "overshooting" and oscillation before a stable level is established. Our results give some evidence that these phenomena occurred during the recovery phase after prestress levels of aspiration had once again been reached. Overshooting was the most prominent of the two; it suggests an analogy between the effects of stress on goal setting behavior in our experiment and the "general adaptation syndrome" described by Selye (1946) after measurements of numerous bodily reactions to quite different forms of stress. The analogy is supported by Berkeley's (1952) results which showed that exposure to the same conditions we employed is associated with physiological changes of the kind observed by Selye. Because we had no way of estimating before the experiment began when in the recovery process overshooting and oscillation might occur, we did not extend the recovery phase sufficiently to observe the predicted final establishment of a stable level.

SUMMARY AND CONCLUSIONS

"Stress" and "goal setting behavior" have been viewed in terms of a model that has been frequently applied to self-regulating physical, biological, and behavioral systems. The model permits certain predictions about the characteristics of goal setting behavior under constant operating conditions and under conditions of stress induced by failure to achieve established goals. The predictions have been tested experimentally using the familiar level of aspiration technique in which feedback, knowledge of results, from the output, performance of a task, affects the behavior system, i.e., establishes or alters levels of aspiration or goals, which in turn influence the future response of the system. Our findings and conclusions may be summarized as follows:

1. The goal setting process very quickly led to the establishment of stated goal levels that remained extremely stable when the conditions under which subjects were operating did not change.

2. The goal setting process was very sensitive to changes in operating conditions. Decreases in stated levels of aspiration were clearly apparent in the very first trials after the introduction of stress; increases in levels were equally apparent immediately following reintroduction of prestress conditions.

TABLE 7

t VALUES[a] OF DIFFERENCES IN RECOVERY TRIALS BETWEEN THE VARIOUS EXPERIMENTAL GROUPS: COMBINED REPLICATIONS

Group	E_1	E_2	E_3	E_4
E_1				
E_2	2.861			
E_3	4.367	1.831		
E_4	3.892	2.099	0.353	

[a] All are significant except the difference between E_3 and E_4 ($df = 38$).

3. Under stress conditions there was a significant trend for stated levels of aspiration to decrease consistently as the number of stress trials increased. The decrease occurred, however, at a decelerating rate, as if the goal setting process were approaching a new steady state level despite continued exposure to stress. Results of other studies suggest that this may be dependent upon the existence of a lower limit—determined by other variables known to affect the behavior system, such as age and sex of the subjects, their socioeconomic status, past success and failure, intelligence—beyond which levels of aspiration will not be set.

4. Reintroduction of prestress conditions was associated with an immediate increase in stated levels of aspiration, which continued at a decelerating rate with increasing numbers of recovery trials until a level was reached that was equal to the prestress level and corresponded to the steady state level maintained by control subjects throughout all phases of the experiment. The results picture the goal setting process as following a similar course under both stress and recovery conditions, while affecting level of aspiration in opposite directions.

5. Increasing numbers of stress trials were associated with increasing numbers of trials to recover prestress levels. Under the conditions of the experiment this trend approached an asymptote between 30 and 40 stress trials.

6. Feedback systems tend to respond to change by "overshooting" and oscillation before a new stable level is established. There was evidence that these phenomena occurred during the recovery phase of the present experiment after prestress levels of aspiration had once again been reached. Overshooting was the most prominent of the two.

7. Although our principal interest was in goal setting behavior, we perforce observed the effects of our experimental treatments on our subjects' levels of performance. We were not able to obtain meaningful measures of performance during the stress phase, performance scores being determined artificially by the experimenter. Analysis of the data we did obtain established that levels of performance showed the same general features as levels of aspiration under the same experimental treatments.

We believe that our results provide new information about the process of goal setting, which is frequently involved in normal human reactions to stress. Goal setting has been viewed in terms of a model which we have found useful for integrating results from previous research and for generating predictions, some of which we have put to test in the present experiment.

REFERENCES

BERKELEY, A. W. Level of aspiration in relation to adrenal cortical activity and the concept of stress. *J. comp. physiol. Psychol.*, 1952, **45**, 443–449.

DAVIS, R. C. The domain of homeostasis. *Psychol. Rev.*, 1958, **65**, 8–13.

FRANK, J. D. Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 119–126. (a)

FRANK, J. D. Some psychological determinants of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 285–293. (b)

HOAGLAND, H. The human adrenal cortex in relation to stressful activities. *J. Aviat. Med.*, 1947, **18**, 450–464.

JONCKHEERE, A. R. A distribution-free *k* test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.

MOSIER, C. A., & HALL, J. R. The social grading of occupations. In D. V. Glass (Ed.), *Social mobility in Britain.* London: Routledge, 1954.

PINCUS, G., & HOAGLAND, H. Steroid excretion and the stress of flying. *J. aviat. Med.*, 1943, **14**, 173–193.

ROTTER, J. B. Level of aspiration as a method of studying personality: I. A critical review of methodology. *Psychol. Rev.*, 1942, **49**, 463–474. (a)

ROTTER, J. B. Level of aspiration as a method of studying personality: II. Development and evaluation of a controlled method. *J. exp. Psychol.*, 1942, **31**, 410–422. (b)

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

RUSSELL, R. W. The comparative study of "conflict" and "experimental neurosis." *Brit. J. Psychol.*, 1950, **41**, 95–108.

RUSSELL, R. W. Behavior under stress. *Int. J. Psycho-Anal.*, 1953, **34**, 1–12.

SELYE, H. The general adaptation syndrome and the diseases of adaptation. *J. clin. Endocrin. Metab.*, 1946, **6**, 117–230.

(Received August 17, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 389-395

RIGIDITY IN NEUROSIS AND GENERAL PERSONALITY FUNCTIONING[1]

LAWRENCE A. PERVIN
Harvard University

THE phenomenon of rigidity has been of interest to psychologists for a number of years. While a vast amount of research on rigidity has been done, our understanding of the manifold aspects of the construct remains at a minimum. One of the difficulties has been the inconsistency of results obtained by various investigators and by any single investigator. A recent review of the literature by Chown (1959) clearly points up this problem.

As regards the generality of rigidity, the question usually asked is: "If a subject is high on rigidity on one task, will he also be high on another?" The Einstellung Water-Jug Problem is frequently used as a comparative measure in such experiments. Cowen, Wiener, and Hess (1953) designed an alphabet task and Wesley (1953) devised a form to measure manifest rigidity. Both found positive correlations with rigidity on the Water-Jug Problem. While Applezweig (1954) concluded that there is no general factor of rigidity, Eriksen and Eisenstein (1953) and Schaie (1955) (using different tasks) found some correlations. Because of the lack of comparability of results, and the contradictory findings reported, the generality of rigidity remains an unresolved problem. The contradictory results suggest that rigidity may not be a general personality characteristic, but may be limited to particular areas of personality functioning. While it is possible that rigid people are rigid in all aspects of their functioning, it is equally possible that their rigidity is specifically related to certain areas of personality functioning. Such a view was in part responsible for the inclusion of various subtests within measures of intelligence. Thus, while this possibility has been noted elsewhere in the past, the tendency has been to look for broad characteristics with few attempts to look into the more complex aspects of the phenomenon. One part of the present study attempts to examine rigidity in different areas of personality functioning.

Another major point of concern in past research on rigidity has been its relationship to personality variables. In abnormal psychology, rigidity is often used to describe behavioral aspects of patients. Mandl (1954), for example, found paranoid schizophrenics to be more rigid than normals. Fey (1951) found normals to be significantly superior to schizophrenics on the Wisconsin Card Sorting Task. Pullen and Stagner (1953) found that those schizophrenics rated as improved after EST showed a significant decrease in rigidity when compared to those who did not improve.

Of the many studies on rigidity done between the years 1948 and 1958, only three seemed to be directly related to neurosis. Using different tasks, investigators reported contradictory findings. In a qualitative analysis of her data, Angyal (1948) found that rigidity on a perceptual task corresponded to certain clinically neurotic manifestations. Fisher (1950) found neurotics to be significantly less rigid than normals on only a few of the tests given, but most of the results were in the predicted direction. Fisher concluded that no single measure on any task was able to measure general personality trends or predict efficiently. Phillipson (1955) found normals to be significantly more rigid than neurotics.

That so few studies should have been done on neurotic rigidity is surprising in view of the great emphasis that has been put on this phenomenon in the functioning of the neurotic. In referring to the neurotic, Fenichel (1945) speaks of the "repetition of action or attitudes once proved useful that occur whenever similar conditions again prevail or generally, repetition in response to repeating stimuli" (p. 342). White (1956), in referring to early neurotic phenomena, talks of the

---

[1] This investigation was carried out during the tenure of a Predoctoral Fellowship (MF-9173) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The author would like to express his gratitude to Herbert C. Kelman of Harvard University and Harold Goodglass of the Boston Veterans Administration Hospital. Their guidance was of invaluable assistance throughout all stages of this study. I would also like to thank Frederick Mosteller of Harvard University for his assistance in the statistical analysis.

> stereotyped, rigid repetition of certain acts in play. They display the earnest, rigid intensity which characterizes neurotic phenomena. The underlying pathology is always a kind of behavior that displays automatic repetition irrespective of the situation, the utility or consequences of the act (p. 237).

From such reports and clinical observations, one may conclude that rigidity is a central phenomenon in the functioning of the neurotic. Therefore, it seems worthwhile to attempt to understand what neurotic rigidity is, and how and why it operates.

The present study examines rigidity in various areas of behavior in neurotics and normals. As far as was possible, an attempt was made to use tasks reported in past research and thereby to provide some basis for comparability of results.

METHOD

Subjects. The subjects were 15 normal, hospitalized patients and 15 neurotic, hospitalized patients taken from the Boston Veterans Administration Hospital.[2] The normals were patients on the general surgical and orthopedic wards. The neurotic population consisted of five obsessive-compulsives, five depressives, and five anxiety reactions. These were based on psychiatric diagnosis. The mean age of the neurotics was 34.1 years, with a range from 23 to 48 years. The mean age of the normals was 33.7 years, with a range from 24 to 45 years. The Vocabulary test of the WAIS was given as a measure of intelligence. The mean IQ of the neurotics was 103.3, with a range from 84 to 129. The mean IQ of the normals was 107.2, with a range from 94 to 123. These means and ranges were considered to be sufficiently similar to justify regarding the normal and neurotic groups as comparable on these two variables.

Materials and Procedure. In deciding upon the nature of the tasks to be given, the following definition of rigidity was used: Rigidity is the inability to shift set or response when the conditions of the situation require it, i.e., when the existing set or response is no longer rewarding. In all, five tasks were given to each of the 30 subjects. All tasks were similar in that they measured rigidity as defined above. Each was designed to measure a different area of personality functioning. The tasks and areas measured were:

AREA	TASK
A. Problem Solving	A. Einstellung Water-Jug Problem
B. Motor	B. Cross-out Letter Test
C. Learning	C. Expectancy-Predictive Card Task
D. Perception	D. Tachistoscope Recognition Test
E. Concept Formation	E. Wisconsin Card Sorting Task (WCST)

The WCST and Einstellung tasks have been described elsewhere (Fey, 1951; Luchins, 1942). In the WCST two shifts were required. The first concept was color, the second form, and the third number.

The perceptual task consisted of a control series in which five color words were serially shown in a tachistoscope, decreasing speeds until the series had been correctly recognized on two consecutive trials. The speed at which the subject reached the criterion was called the control series recognition speed. For the experimental series, the exposure times were decreased to the original speed of the control series. Here, for every third color word in the series, a noncolor word was substituted. At any one exposure speed, each noncolor word was once substituted for a color word so that, in all, 15 words were shown (five noncolor words and each of the five color words twice). This was done at each exposure speed until two successive correct recognitions of the entire series. All words were equally frequent in usage according to the Thorndike-Lorge Word Frequency List. The words used were brown-bring, white-where, green-great, black-blood, blue-blow. In the perceptual task, rigidity could express itself in two ways: (*a*) in the increase in number of trials until recognition of the mixed color and noncolor series over the number of trials until recognition of the control series and (*b*) in the number of noncolor words seen as color words after the recognition speed of the control (all color) series had been reached or by the time all words should be correctly recognized. If they were not, the subjects could be said to be operating in terms of a set for color words.

In the Expectancy-Predictive Card Task, the subjects were presented with 120 cards. Each card was either black or red. Half the first 30 cards were black and half red, randomly arranged. In the next 60, the ratio of blacks to reds was 3 to 1, again randomly arranged. The last 30 cards were all red. Subjects were asked to predict the color of each card before turning it over. Three measures of rigidity were used in this task:

1. The percentage called black within Cards 36 through 45 and the percentage called black within Cards 46 through 60 were computed for each subject. The difference between these two percentages was then determined for each subject. The greater the difference, representing an increase in percentage called black, the less the rigidity. Thus, rigidity was manifested to the extent that the subject failed to increase the proportion of cards predicted to be black.

---

[2] I would like to thank the members of the staff and patients at this hospital for their cooperation.

2. The percentage called black within Cards 76 through 90 and the percentage of cards called black within Cards 91 through 105 (all red) were computed for each subject. The difference between these two was determined for each subject. The greater the difference, representing a decrease in percentage called black, the less the rigidity. Thus, rigidity was manifested to the extent that the subject failed to decrease the proportion of cards predicted to be black.

3. In both of the above measures a greater difference represented less rigidity. An overall index of rigidity for this task was computed by combining the above two differences for each subject.

In the motor test, the subjects were asked to cross out every second capital letter and every third "small" letter. The first two lines were all capitals; thereafter, the series of capitals and lower-case letters were intermixed. An error in this test was either a letter not crossed out that should have been or one crossed out that should not have been. The number of errors was taken as the measure of rigidity.

These five tests were given to each subject in a single experimental session. To control for the effect of each task upon the subsequent tasks, the orders of tasks were counterbalanced.

RESULTS

The results obtained from the rigidity battery were first evaluated by comparing the scores of the normal group with those of the neurotic group, for each of the individual measures of rigidity. Single-tailed t tests are used throughout to determine group differences. A summary of the results of the major measures of rigidity, to be discussed below, is presented in Table 1.

TABLE 1

SUMMARY OF THE DIFFERENCES IN RIGIDITY BETWEEN NEUROTICS AND NORMALS IN VARIOUS AREAS OF PERSONALITY FUNCTIONING

Area	Normal Mean	Neurotic Mean	t	p Value[a]
Problem Solving	2.40	1.86	.57	<.30
Motor	2.73	4.80	1.75	<.05
Learning				
Difference between Cards 31–45 and 46–60	.10	.03	1.23	<.15
Difference between Cards 76–90 and 96–105	.37	.19	2.52	<.01
Combined Difference	.48	.15	3.46	<.01
Perception				
Trials	5.31	6.46	2.72	<.01
Words	2.42	4.27	1.55	<.10
Concept Formation	5.47	29.47	5.11	<.001

---

[a] Single-tailed test of significance.

Problem Solving. The mean number of problems solved by the short method were 1.86 for neurotics and 2.40 for normals. The difference was in the predicted direction, but the p value was less than .30. Eight neurotics and seven normals were unable to solve the problem that necessitated the short-cut method. Neurotics were slower on the criterion problems, but since they were slower on all others as well, time was not a discriminating factor between the two groups.

Motor. The mean number of errors were 4.8 for neurotics and 2.7 for normals. The difference in means was found to be significant at the .05 level of confidence. Since time could have influenced accuracy, the times for the first two lines (all capitals) and whole test were recorded. The greater the time spent on the first two lines, the less the proportion of time used for the rest of the test. This would represent less difficulty in the shifts that came after the first two lines and consequently less rigidity. The mean percentage of time for the first two lines were .249 for neurotics and .260 for normals. Thus, the difference in errors was not due to an emphasis on accuracy at the expense of time by the normals. (Normals did better, but not because they took more time.)

Learning. Using the first measure of rigidity, the mean difference in the two percentages was found to be .027 for neurotics and .094 for normals. The resulting p value was between .10 and .15.

Using the second measure of rigidity, the mean difference in the two percentages was .192 for neurotics and .367 for normals. The difference was found to be significant at the .01 level of confidence.

Using the combined index, the mean combined difference (for the two above measures for each subject) was .148 for neurotics and .477 for normals. The difference was found to be significant at beyond the .01 level of confidence.

The greater difficulty in shifting their probability bases on the part of the neurotics can be seen from the data given in Table 2.

Perception. The mean number of trials until correct recognition of the series of color words were 3.5 for neurotics and 3.6 for normals. Thus, there was apparently no difference in perceptual ability per se. The

TABLE 2

DIFFERENCES IN PERFORMANCE AMONG NEUROTICS AND NORMALS ON THE EXPECTANCY-PREDICTIVE CARD TASK

Cards	Percentage called black (neurotics)	Difference	Percentage called black (normals)	Difference	p Value[a]
31–45	.462	.026	.492	.095	>.10 <.15
46–60	.488		.587		
61–75	.579		.592		
76–90	.591	.192	.675	.367	<.01
91–105	.399		.308		
106–120	.023		.031		

[a] Single-tailed test of significance.

mean number of trials for correct recognition of the mixed series (color and noncolor) were 6.4 for neurotics and 5.2 for normals. This difference was significant at the .01 level of confidence. Thus, the neurotics had more difficulty in correctly recognizing the words of the mixed series than did normals. To what was this due? On the second measure of rigidity in this task, for neurotics, the mean number of noncolor words called color once the control series recognition speed had been reached was 4.2. The corresponding mean for the normals was 2.4. This difference represents a p value between .05 and .10. This difference between the groups cannot be accounted for on the basis of some "guessing tendency" because until the control recognition speed had been reached, normals made more such errors than did neurotics.

Concept Formation. On the WCST, the mean number of trials until criterion for recognition of the color concept were 5.53 for neurotics and 5.27 for normals. Thus, at this point there would seem to be little difference in what has been called concept formation ability. The first change required was that from color to form. The mean number of trials until criterion for recognition of the form concept were 9.93 for neurotics and 6.87 for normals. This difference represents a p value between .20 and .25. The second change that was required was that from form to number. The mean number of trials until criterion for recognition of the number concept were 29.47 for neurotics and 5.47 for normals. This difference was significant at the .001 level of confidence. When the nature of the responses is further analyzed, it is seen that the neurotic mean for perseverative responses is 27.2 while the corresponding mean for normals is 5.0.

Intercorrelations of Tasks. Following the above analysis the intercorrelations of the various measures of rigidity were studied for both normal and neurotic populations. The Pearson product-moment correlation coefficients are represented in Table 3. As can be seen, the correlation coefficients are generally low. Out of the 20 correlation coefficients only 1 reached significance, about what would be expected by chance. Furthermore, the correlation coefficients are different for the two groups. The correlations of the tests with age and IQ were also computed. These correlation coefficients were generally low and somewhat inconsistent.

In the initial conceptualization of the problem it was suggested that each test had something in common with the other tests, which may be called general rigidity, but also had a rigidity specific to it, depending upon the area being represented within the test. It was therefore decided to try to get an estimate of the rigidity in each test that is common to the rigidity in the other tests. In other words, to find out how much of each test is common to all of the other tests. This analysis was also used in an attempt to explain the inconsistency of the correlations noted in Table 3. If, as in Table 3, each test is correlated with any one test, the uniqueness of the two tests is emphasized. However, if each

TABLE 3

PRODUCT-MOMENT CORRELATION COEFFICIENTS OF TESTS OF RIGIDITY FOR NORMAL AND NEUROTIC POPULATIONS

	WCST	Cross-out Letter Test	Water-Jug Problem	Word Recognition Test
Normals				
E-P Card Task	.10	−.01	.24	.14
WCST		.16	.59	.11
Cross-out Letter Test			.16	.29
Water-Jug Problem				.18
Neurotics				
E-P Card Task	−.09	−.14	−.32	−.32
WCST		.44	.22	.35
Cross-out Letter Test			.33	−.02
Water-Jug Problem				.24

test is correlated with the sum of the other four tests, the rigidity common to the set is emphasized. With the latter measure less emphasis is put upon the particular characteristics of the population involved, i.e., in which areas the subjects are rigid. It was hoped that by doing this the correlation coefficients would remain fairly stable from group to group. As a reasonable estimate of what the tests have in common, each test was correlated with the sum of the other tests by assuming equal weights and using the following formula (for Test 5, for example):

$$r_{y5} \sum_{i}^{4} yi$$

$$= \frac{r_{15} + r_{25} + r_{35} + r_{45}}{\sqrt{4 + 2(r_{12} + r_{13} + r_{14} + r_{23} + r_{24} + r_{34})}}$$

For each of the tests the rigidity measured may be different, but it is felt that this difference is slight, i.e., there is enough in common to call them by the same term. The reasoning behind this is: Each task is correlated with the sum of the other four. For each of the five tasks the composition of the sum with which each is correlated is slightly different. But, three of the four tasks composing the sum will be the same for any two correlations. To assess what is exactly common to the set of five variables requires a more extensive analysis than is given here.

The correlation coefficients for each test of rigidity with the sum of the other four tests of rigidity are given in Table 4. The correlation coefficients in Table 4 indicate that each test measures, in part, a factor common to the other tests. While the correlation coefficients for the normal and neurotic groups are fairly similar, they are not as similar as would have been expected. An examination of the correlations presented in Table 3 indicates that this is mainly because the correlations of the Expectancy-Predictive Card Task with the other tasks are different for normals and neurotics. This task correlated negatively with the other tasks for neurotics and predominantly positively with the other tasks for normals. If this task is eliminated and each of the tests is correlated with the sum of the other three, the resulting coefficients are as given in Table 5. As can be seen, the resulting correlation coefficients are similar for the two groups. This would indicate that the four tasks have something in common and there are no large differences in what they mean for normals and neurotics.

TABLE 4

CORRELATIONS OF EACH TEST OF RIGIDITY WITH THE SUM OF THE OTHER FOUR TESTS OF RIGIDITY FOR NORMAL AND NEUROTIC POPULATIONS

	Normals	Neurotics
E-P Card Task	.18	−.12
Water-Jug Problem	.49	.22
WCST	.43	.25
Cross-out Letter Test	.23	.29
Word Recognition Test	.11	.11

TABLE 5

CORRELATIONS OF EACH TEST OF RIGIDITY, EXCLUDING THE E-P CARD TASK, WITH THE SUM OF THE OTHER THREE TESTS OF RIGIDITY FOR NORMAL AND NEUROTIC POPULATIONS

	Normals	Neurotics
Water-Jug Problem	.39	.33
WCST	.37	.44
Cross-out Letter Test	.25	.31
Word Recognition Test	.24	.23

DISCUSSION

The results presented above would appear to be consistent with and support the clinical observations that rigidity is a characteristic of neurosis. It would seem that the nature of the control population would tend to further emphasize the differences. These patients were also ill and hospitalized. Furthermore, the mental health of some of the "normal" patients could be questioned. One of the control patients said that he had been discharged from the army because of bad legs, but had received a Rorschach before leaving. Two of the patients in the control group had ulcer conditions. Also, many of the neurotic patients were discharged shortly after testing. Thus, the sampling problem may have led to a reduction of differences.

While four of the five tests resulted in significant differences, the difference on the number category of the WCST is probably the most dramatic. A number of possible explanations for this result may be given: it may be that the number concept per se is particularly difficult for neurotics; it may be that because a second change is required, the number concept was particularly difficult for neurotics. While other explanations are possible, those given can be studied by rearranging the order of concepts so that number is either the first or second concept. It may be, however, that rigidity becomes particularly evident in a difficult, challenging, or ambiguous situation. Thus, perhaps it would also be of value to use another difficult concept. This could be done by extending the WCST to include another concept (such as shading) or by requiring a concept such as color and form together (allowing some not to be matched at all).

In view of the clear differences found between the two groups, how might the role of rigidity in neurosis be conceptualized? If it may be granted that humans strive for control of their environment, that control is partly dependent upon knowing what to expect, and that one aspect of new and ambiguous situations is that in them one does not know what to expect, then rigidity in neurosis may be seen as operating as follows: Like all humans, the neurotic strives for control of his environment. However, the neurotic has less control over his environment than does the normal person, and experiences his accordingly. Therefore, situations which call into question his ability to master the environment are a particular problem for the neurotic. New and ambiguous situations create such a problem. While normals are usually able to confront such situations and attempt to develop new patterns of behavior appropriate to them, the neurotic flees from the situation in which his ability to control and master is challenged. He interprets the new in terms of the old and therefore feels that old patterns of behavior are still appropriate and adequate. As in the Expectancy-Predictive Card Task, he has difficulty in changing his expectancies. By expecting the future to be similar to the past, he feels justified and adequate in using old patterns of behavior. Rigidity is useful to the neurotic in that it assists him in coping with new and ambiguous situations which would otherwise challenge his doubtful ability to cope with the environment.

It has been demonstrated that the correlations between different tasks measuring rigidity are generally low and vary from group to group. Rigidity is not a general personality characteristic. Within the framework of the present study, the results indicate that a person may be rigid in one area of personality functioning and not in another. If this is true, then it vastly complicates our attempts to study personality traits and characteristics.

The analysis of the correlations for each task with the sum of the other tasks suggested that the Expectancy-Predictive Card Task measures something different for the two groups. For one group this test could be measuring rigidity in the sense of not attending to new information, while for the other group it could be measuring rigidity in the sense of not changing responses even though new information has been received. At this point, no definitive answer to this problem can be given.

Finally, the results indicate that a neurotic may be more rigid than a normal in one or more areas of personality functioning, but not necessarily in all. This suggests the possibility that the particular area or areas within which the neurotic is rigid is of some significance. The results within the neurotic population suggest that different kinds of neurotics (obsessive-compulsive, anxiety, depressive) are high in rigidity on different tasks. While the number of cases within each group (five) is too small to report confidently on these differences, some trends may be noted. There were indications that the anxiety neurotics were particularly rigid on the motor tasks, while the obsessive-compulsives were comparatively more rigid on the more cognitive tasks. The particular areas in which a group of neurotics are rigid may be peculiarly relevant to the nature of their problem areas or ways of handling their problems. It would appear that further investigations into within-neurotic differences might prove helpful in understanding the disorders.

SUMMARY

The present investigation reported on rigidity in various areas of behavior in neurotics and normals. Each of the 15 hospitalized normals and 15 hospitalized neurotics took five tests of rigidity. Each test was designed

to test rigidity in one of the five following areas: problem solving, motor, learning, perception, and concept formation. Neurotics were found to be more rigid than normals on each of the five tests. Four of the five tests of rigidity resulted in differences between the two groups which were significant at the .05 level of confidence or better.

The intercorrelations of the tests of rigidity were generally low for both the normal and neurotic populations. It was suggested that individuals may be rigid in one area of personality functioning and not in another.

A partial theory of rigidity in neurosis was given along with some suggestions for further research.

REFERENCES

ANGYAL, ALICE F. The diagnosis of neurotic traits by means of a new perceptual test. *J. Psychol.*, 1948, **25**, 105–135.

APPLEZWEIG, DEE G. Some determinants of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1954, **48**, 224–228.

CHOWN, SHEILA M. Rigidity—a flexible concept. *Psychol. Bull.*, 1959, **56**, 195–223.

COWEN, E. L., WIENER, M., & HESS, J. Generalization of problem-solving rigidity. *J. consult. Psychol.*, 1953, **17**, 100–103.

ERIKSEN, C. W., & EISENSTEIN, D. Personality rigidity and the Rorschach. *J. Pers.*, 1953, **21**, 386–391.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FEY, ELIZABETH T. Performance of young schizophrenics and young normals on the WCST. *J. consult. Psychol.*, 1951, **15**, 311–319.

FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**(1, Whole No. 307).

LUCHINS, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, **54**(6, Whole No. 248).

MANDL, BILLIE SUE T. An investigation of rigidity in paranoid schizophrenics as manifested in a perceptual task. Unpublished doctoral dissertation, Purdue University, 1954.

PHILLIPSON, L. K. Personality ridigity as a distinguishing factor between normal and neurotic behavior. Unpublished doctoral dissertation, Purdue University, 1955.

PULLEN, M. S., & STAGNER, R. Rigidity and shock therapy of psychotics. *J. consult. Psychol.*, 1953, **17**, 79–86.

SCHAIE, K. W. A test of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1955, **51**, 604–610.

WESLEY, ELIZABETH. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–143.

WHITE, R. W. *The abnormal personality.* New York: Ronald, 1956.

(Received August 18, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 396–401

TWO CORRELATES OF SYMBOLIC SEXUAL AROUSAL: LEVEL OF ANXIETY AND LIKING FOR ESTHETIC MATERIAL[1]

MICHAEL A. WALLACH
Massachusetts Institute of Technology

THE present research tests two hypotheses derived from Freudian theory concerning the symbolic expression of sexual arousal. The first holds that symbolic expression of sexual arousal is required to a greater extent in persons whose level of anxiety is high, since such individuals are more likely to inhibit the direct or manifest expression of sexual arousal. This hypothesis follows from Freud's writings on dreams (1953a). Relationships between symbolic and manifest projections of sexual arousal have been studied by Clark and Sensibar (1958), and evidence that symbolic sexual arousal in fantasy production permits the expression of impulses denied an overt outlet has been presented in Wallach and Greenberg (1960).

The second hypothesis to be tested states that degree of liking for esthetic materials correlates positively with the degree to which symbolic expression of sexual arousal occurs in response to the materials. This prediction follows from Freud's (1953b) conception of art as a "sublimation" of sexual energy. Proposals toward the study of motivational variables in esthetics have been made by Wallach (1959), and an experiment in this direction concerning the motivational functions served by aspects of graphic expression has been reported by Wallach and Gahm (1960).

While a relationship between degree of liking and degree of symbolic sexual arousal was not found for jazz music in the Wallach and Greenberg study, it seemed very possible that jazz was an inappropriate medium for testing the hypothesis, because degree of preference for jazz may primarily reflect determinants other than an evaluation of the music as such. Preferences regarding jazz may be importantly influenced by judgments of jazz associated activities and things, such as dancing, drinking, and night clubs. It therefore seemed advisable to obtain preference judgments that might reflect evaluation of the music as such more than was the case with jazz. In the present study, degree of liking for certain "classical" selections—music in the Debussy-Ravel idiom—was used for this purpose.[2]

METHOD

Experimental Subjects

The subjects were undergraduate women in introductory psychology classes at Simmons College, 73 of whom participated in the story writing and preference rating sessions. Sixty of these subjects also took part in the session at which anxiety was assessed. Sessions of the experiment were administered during class periods without prior announcement. Subjects identified themselves by name on all procedures.

Experimental Sessions

Story session. The first class period was devoted to the presentation of three musical selections, with the request to write a story suggested by the piece of music just heard. The procedure was described as a test of creative imagination, and subjects were asked to let the mood induced by the music determine the story that they wrote. A 4-minute period for story writing was provided after the playing of a selection, the subject being permitted to write notes during the music if desired, but not to start the story itself until the music had ended. A blank sheet was in front of the subject while listening to each selection, and each story was written on a separate page. These stories provided the raw material from which symbolic sexual arousal was assessed. In this and all other sessions, a male experimenter presided, introduced as a professor from another university.

Preference session. During a second class period, about a month later, the same subjects were asked to make a judgment of degree of liking for each of the three pieces. After hearing each selection again in the same order used in the story session, they were to

---

[1] This investigation was supported by a research grant (M-2269) from the National Institute of Mental Health, United States Public Health Service, conducted under the auspices of the Age Center of New England, Inc. Grateful thanks are due Carol Greenberg, Michael Humphreys, Kermit Roosevelt, and Charles Stewart, for aid in data analysis; Donald S. Dunbar and Thomas F. Pettigrew, for making classes available; and Roger Brown, Nathan Kogan, and Lise Wallach, for critical comments on the manuscript.

[2] The term "classical" is used here in a nontechnical sense as shorthand for referring to the Debussy-Ravel selections presented, in contrast to the jazz selections of the previous study by Wallach and Greenberg (1960).

circle a number on a seven-point scale labeled "strongly dislike" at 1 and "strongly like" at 7. It was emphasized that there were no right or wrong answers.

Anxiety assessment session. Level of anxiety was assessed from the Neuroticism scale of the Maudsley Personality Inventory (Eysenck, 1956; Jensen, 1958). Jensen notes that for American university students, the Neuroticism scale correlates .77 with the Taylor Manifest Anxiety Scale (Taylor, 1953). In addition, social extraversion-introversion was assessed from the Extraversion scale of the same instrument, which has a correlation of .81 (Jensen, 1958) with the Social Extraversion scale of the Minnesota TSE Inventory (Evans & McConnell, 1957). The whole untitled questionnaire was filled out during a class session occurring about a month after the musical preference ratings.

Content Analysis of the Stories for Symbolic and Manifest Sexual Arousal

The categories used for scoring the stories in terms of symbolic sexual arousal have been described in detail in Wallach and Greenberg (1960). These categories were originated by Clark and Sensibar (1958) and Beardslee and Fogelson (1958). They include five variables, each of which is scored for presence or absence in a story: motion (description of movement by the body as a whole or by any part of the body from the shoulders down); rhythm (description of any repetitive action, sound, or adverb phrase); peak (a rise followed by a decline in the level of tension implied by the action); penetration (imagery concerning forceful pushing through a resistant medium); and symbolic content (presence of one or more items of story content which are listed by Gutheil—1951, Ch. 3—as sexual symbols, in his formalization of Freud's—1953a—theory of dreams). Because of its direct associations with music, imagery concerning dancing was not scored.

The interpretation that these variables concern symbolic sexual arousal rests on several grounds. First, Clark and Sensibar (1958) have shown that male subjects sexually stimulated by pictures of nude women project more "symbolic content" (the fifth category defined above) than nonstimulated males, both groups being tested in a classroom situation. Second, Wallach and Greenberg (1960) found that the frequency of occurrence of "symbolic content" tends to covary with the frequency of occurrence of the remaining four categories, and the same result was obtained in the present study. Since the five variables cluster empirically, there would seem to be support for Freud's view that all five are symbolic of sexual arousal. Third, an independent sample of female judges was asked to rate the sexual arousal values of the musical selections used in the present study. It will be seen that these independent ratings of the selections for sexual arousal value closely match the average symbolic sexual arousal scores from stories written to the various selections; the same kind of correspondence also was obtained in the previous study by Wallach and Greenberg.

Besides being scored for symbolic sexual arousal, the stories also were scored for manifest sexual arousal according to the system used by Clark (1955). Each of the following kinds of manifest sexual imagery was scored for presence or absence in a story: primary (some kind of evidence for sexual intercourse); secondary (reference to secondary sexual activity such as kissing); tertiary (reference to courting, dating, etc.). As previously noted, dancing imagery was not scored, contrary to Clark's practice.

Two independent scorers scored the 219 stories of our 73 subjects. As each story appeared on a separate sheet of paper identified only by a randomly assigned number, the scorers worked "blindly," not knowing which stories were written by the same subject, nor which stories were written to a particular piece of music.

Prior Selection of "Classical" Pieces and Independent Judgments of Arousal Value

An introductory psychology class of 29 undergraduate women from Radcliffe College heard the three musical selections, and were instructed to rate each piece on a seven-point scale, from "least arousing" (1) to "most arousing" (7), after hearing the selection, in terms of the degree to which it aroused them sexually. The purposes of this judgment session were to obtain independent explicit evaluations of the cue value of each piece for sexual arousal, and to help set the order in which the three selections might be presented as stimuli for stories to the experimental subjects. It was desirable to present the selections in an order of increasing sexual arousal value, to avoid the likelihood that story imagery stimulated by a high arousal piece would be transferred to a subsequently presented selection lower in arousal value. The tests reported below indicated that one of the three selections was judged less arousing than the others; this selection was therefore presented first to the experimental sample.[3]

The first selection was a passage from the third movement of Ravel's *String Quartet in F Major* (Piece A); the second was a passage from the first movement of the same work (Piece B); and the third was a section from the second movement of Debussy's *Petite Suite* (Piece C). The sections to be presented were taped from records, and their respective lengths were 180, 173, and 125 seconds.[4]

RESULTS AND DISCUSSION

Preliminary Considerations

Independent judgments of sexual arousal values. The independent judgments of arousal value made by the 29 Radcliffe women yielded the following means (higher numbers signifying higher arousal value): 3.24 for Piece A, 4.07 for B, and 4.10 for C. The judged arousal value of Piece A thus seems considerably lower than that for B and for C. The difference in a

---

[3] This same order of presentation had been used for the independent judges.

[4] The Ravel recording was Columbia ML 5245; and the Debussy, Epic LC 3165.

TABLE 1
MEAN SYMBOLIC SEXUAL AROUSAL TO "CLASSICAL" MUSIC FOR THE EXPERIMENTAL SAMPLE, BY SELECTION AND SCORING CATEGORY
(N = 73)

Selection	Mean symbolic sexual arousal score			Test of mean difference between scores					
	A	B	C	B-A		C-A		B-C	
				t	$p<$	t	$p<$	t	$p<$
Motion	0.25	0.51	0.52	3.48	.01	4.16	.01	0.00	*ns*
Rhythm	0.05	0.18	0.14	2.50	.02	1.59	*ns*	0.84	*ns*
Peak	0.05	0.19	0.08	1.87	.10	0.62	*ns*	1.69	.10
Penetration	0.00	0.03	0.00	0.14	*ns*	0.00	*ns*	0.55	*ns*
Symbolic content	0.41	0.64	0.55	5.66	.01	2.09	.05	0.53	*ns*
Total	0.76	1.55	1.29	5.67	.01	3.70	.01	1.67	.10

judge's evaluation for each pair of pieces was determined, and the mean difference across judges for each pair of pieces was tested for significance of deviation from zero. Comparing A and B, the mean difference yields a t of 2.68 ($p < .02$);[5] comparing A and C, t is 2.13 ($p < .05$); and comparing B and C, t does not approach significance.

Reliability of the scoring system for sexual arousal. Reliability between the two independent scorers was determined for each of the five scoring categories for symbolic sexual arousal. Since a scorer made 219 decisions of presence or absence in the case of each scoring variable, the reliability measure for each variable was taken as the number of stories for which both scorers agreed in their decision of presence or absence, divided by 219. Expressed as percentages, the values obtained were: 87% for motion, 89% for rhythm, 92% for peak, 97% for penetration, and 80% for symbolic content. In a similar manner, reliabilities were determined for each of the three scoring categories for manifest sexual arousal. The number of agreements divided by 219 was 100% for primary, 100% for secondary, and 97% for tertiary sexual imagery. The definitions of the scoring categories hence were clear enough to yield high agreement. Those cases in which the two independent scorers disagreed were resolved by the decisions of a third independent scorer.

Cue value of the selections for sexual arousal. Table 1 presents the mean level of symbolic sexual arousal to each of the three pieces taken separately. Since the same person wrote stories to all three selections, significance tests comparing any two pieces were based on the difference between a person's symbolic arousal scores to those two pieces. On all symbolic arousal scoring categories combined, and for motion, rhythm, and symbolic content taken separately (with a trend for peak in addition), Pieces B and A differ significantly, Piece B being higher. On all categories combined, and for motion and symbolic content taken separately (with a trend for rhythm in addition), Pieces C and A differ significantly in symbolic arousal, Piece C being higher. No differences reach the .05 level for C vs. B. To determine whether these results might depend on story length, three t tests were computed, each evaluating the mean differences in length (number of words) for stories to two pieces: A vs. B, A vs. C, and B vs. C. In each test, the score was the difference in story length for a subject's stories to the two selections in question. None of the t values approaches significance. Thus, symbolic arousal to different pieces varies while the length of stories written to these different pieces does not appreciably change.

As was also found in the Wallach-Greenberg study using jazz selections, there is thus a high degree of consistency between the comparative sexual arousal values of the "classical" selections as rated explicitly by a group of independent judges, and symbolic sexual arousal levels to these same selections as measured by a content analysis of stories written without awareness of the purpose. It might further be mentioned that no cumulative series effect seems to have been operating with regard to symbolic sexual arousal, since symbolic arousal for the last piece was lower (though not significantly so) than that to the preceding one.

---

[5] All significance levels cited in this study are for two-tailed tests.

Turning now to manifest sexual imagery in response to the "classical" pieces, the frequency of occurrence of such imagery was very low. While the average amount of *symbolic* sexual arousal per person for all three pieces and all five categories combined was 3.60, the average amount of *manifest* sexual arousal per person for all three pieces and all three categories combined was 0.70: 0.00 primary, 0.08 secondary, and 0.62 tertiary imagery per person for all three pieces. The amounts of each of the three kinds of manifest imagery to the three musical selections did not differ significantly. In sum, the present experimental situation may not have been suitable for eliciting manifest sexual imagery, perhaps because of the constraint of the classroom and the presence of a male professor.

Anxiety Level and Symbolic Sexual Arousal

With respect to the first hypothesis—that degree of symbolic sexual arousal varies directly with level of anxiety—a 2 × 2 analysis of variance was performed using anxiety level and preference level as treatment variables, and overall symbolic sexual arousal as the dependent variable (see Table 2). The symbolic sexual arousal score for each subject in this and all subsequent analyses consisted of the subject's total score for all five symbolic arousal categories combined, across all three musical pieces combined. Treatment groups were constructed by separating the sample into four cells consisting of high-high, high-low, low-high, and low-low combinations on the anxiety and preference variables. The sample was dichotomized as close to the median as possible for each variable.[6] Since the number of subjects in the four cells varied, Snedecor's method of unweighted means was used (Grosslee, 1959; Snedecor, 1956, pp. 385–388). The four variances were found homogeneous by Bartlett's test (Walker & Lev, 1953, pp. 193–194), and the fixed constants model (McNemar, 1955, pp. 304–306) was used in calculating F ratios.

TABLE 2

MEAN SYMBOLIC SEXUAL AROUSAL TO "CLASSICAL" MUSIC AND ANALYSIS OF VARIANCE, FOR GROUPS HIGH AND LOW IN ANXIETY AND IN PREFERENCE FOR THAT MUSIC

($N = 60$)

	High anxiety	Low anxiety
High preference	4.78	4.07
Low preference	3.95	1.86

Source	*df*	*MS*	*F*
Anxiety	1	26.13	7.62**
Preference	1	30.93	9.02**
Interaction	1	6.40	1.87
Within cells	56	3.43	

** $p < .01$.

The analysis of variance indicates that degree of anxiety is significantly related to level of symbolic sexual arousal ($p < .01$). To further explore this relationship, the product-moment correlation was computed between level of anxiety and symbolic sexual arousal in the stories, yielding an r of .30 ($p < .03$, $N = 60$). The corresponding partial correlation, holding story length constant, is .28 ($p < .04$), indicating that the original relationship is not appreciably influenced by story length.

A possible artifact involved in the anxiety questionnaire measure concerns a tendency to agree in one's answers to any sorts of questions (Couch & Keniston, 1960). Two lines of evidence suggest that such a "yea saying" tendency is not a significant factor in the present study. Since all the items in the anxiety scale are keyed so that "yes" responses lead to a high anxiety score, and two-thirds of the items in the social extraversion scale, which also was administered, are similarly keyed with respect to extraversion, an agreeing response set would lead to a significant positive correlation between the two scales. The obtained correlation, however, is not significant and is actually negative. Further, if yea saying were responsible for the obtained relationship between anxiety and symbolic arousal, it also should result in at least a moderate relationship between extraversion and symbolic arousal, since the latter scale resembles the former in having the majority of its items keyed so that "yes" responses yield a high score. The relationship between extraversion and symbolic arousal, however, did not even approach significance ($r = -.01$).

---

[6] The splits made were 52%–48% for anxiety, and 60%–40% for preference.

Not only an anxiety effect but also a significant interaction between anxiety level and social extraversion-introversion was obtained in the previous study of symbolic arousal to jazz: anxious introverts showed the highest symbolic arousal, nonanxious introverts the lowest, and extraverts fell between these extremes. The question arises as to why no comparable interaction was obtained in the present study. Perhaps the answer is that jazz, because of the social activities and things typically associated with it, has greater relevance to social extraversion-introversion than does the more "classical" music used here. Some support for this suggestion comes from the finding of a significant relationship between degree of social extraversion and degree of overall preference for the jazz ($r = .33$, $N = 72$, $p < .01$), while no such relationship was obtained between social extraversion and degree of overall preference for the "classical" selections.

Finally, no relationship was found between anxiety and manifest sexual imagery (nor between preference for the "classical" pieces and manifest sexual imagery). It has already been noted, however, that the overall amount of manifest sexual imagery in the present experimental situation was very low.

In sum, a relationship has been demonstrated between anxiety level and symbolic sexual arousal, which confirms the first hypothesis of the study. The causal sequences here, however, need further clarification. On the one hand, the subjects may be equally aroused sexually, and those who are more anxious may be constrained to express their arousal symbolically rather than directly—the initial Freudian view. On the other hand, it is possible that more anxious subjects resort more to sexual symbolism because their overall sexual sensitivity or arousability is greater, their stronger sexual arousability making them more anxious and hence more likely to express their arousal only at a symbolic level. Subjects low in anxiety, in turn, may lack anxiety precisely because their overall sexual arousability is low.

Liking for Esthetic Material and Symbolic Sexual Arousal

The second hypothesis stated that degree of symbolic sexual arousal in response to the "classical" selections varies directly with overall liking for the selections. The preference measure was obtained by summing a subject's preference ratings for the three "classical" pieces.

The analysis of variance in Table 2 indicates that degree of preference for the "classical" selections covaries with symbolic arousal. The preference effect, like the anxiety effect, is significant beyond the .01 level. In addition, preference level for the "classical" pieces correlates significantly with degree of symbolic sexual arousal to them ($r = .32$, $p < .01$, $N = 73$). The partial correlation between these two variables, holding story length constant, is .27 ($p < .03$), indicating no appreciable influence by story length differences. That this relationship is not due to a response set to be extreme in marking rating scales may be seen from an r of .01 ($N = 67$) between the sum of a person's preference scale ratings for the "classical" pieces and for the jazz pieces.

Whether symbolic sexual imagery in the stories reflects different basal levels of arousability apart from the musical stimulus or reflects the specific music heard was not at issue with respect to the relationship between anxiety level and symbolic sexual arousal. It becomes relevant, however, in connection with testing the present hypothesis of a relationship between preference for the "classical" pieces and symbolic sexual arousal to them. It is possible that the significant correlation found between these two variables might also be obtained between preference for the "classical" pieces and symbolic sexual arousal to different stimuli. In fact, no correlation is found, however, between preference for the "classical" pieces and symbolic sexual arousal to *jazz* pieces ($r = -.04$), suggesting that the obtained relationship between preference and symbolic arousal for the "classical" pieces at least partly reflects this "classical" music as such.

As a further test of a relationship between preference and symbolic arousal, the mean preferences for the three pieces were inspected. Since mean symbolic arousal was lower for A than for B and C, average preference levels might be expected to follow the same trend. Mean degrees of preference (higher numbers signifying greater preference) are 4.64 for Piece A, 5.10 for B, and 6.33 for C. The difference in a judge's evaluation for each pair of pieces was obtained ($N = 73$ in each case). Comparing A and B, the mean difference has a t of 2.04 ($p < .05$); for A and C, t is 3.75 ($p < .01$); and for B and C, the t is 6.30 ($p < .01$). Piece A thus was liked less than B or C. Degree of arousal to and degree of preference for the various selections hence are related. Although both B and C were liked more than A, however, C also was liked more than B. Thus, factors besides symbolic sexual arousal level undoubtedly influence preference.

The basis of esthetic preference is a central question for the philosophy of esthetics, and philosophers (e.g., Hospers, 1946) have proposed various criteria in trying to understand how people assess esthetic worth. Freud's "sublimation" theory would predict a positive relationship between degree of liking for a work of art and the degree to which the work permits symbolic expression of sexual energy, since sublimated expression is supposed to be the main psychological function of art. While symbolic sexual arousal obviously is not the sole correlate of esthetic preferences, the results of the present study are in keeping with the sublimation hypothesis. We cannot yet, of course, be sure that the causal sequence is, as Freud would suggest, from symbolic sexual arousal to preference.

SUMMARY

The present study tests two Freudian hypotheses concerning the symbolic expression of sexual arousal: that symbolic sexual arousal occurs to a greater extent in persons of high anxiety level, and that degree of liking for esthetic materials is positively related to the degree of symbolic sexual arousal in response to those materials. Both hypotheses were confirmed in an experiment with 73 female undergraduates involving the content analysis of imaginative stories written in response to music. Possible implications for anxiety and for the psychology of art were suggested.

REFERENCES

BEARDSLEE, D. C., & FOGELSON, R. Sex differences in sexual imagery aroused by musical stimulation. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 132–142.

CLARK, R. A. The effects of sexual motivation on phantasy. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century, 1955. Pp. 44–57.

CLARK, R. A., & SENSIBAR, MINDA R. The relationship between symbolic and manifest projections of sexuality with some incidental correlates. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 117–129.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

EVANS, CATHERINE, & MCCONNELL, T. R. *Minnesota T-S-E inventory*. (Rev. ed.) Princeton: Educational Testing Service, 1957.

EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, 1956, **54**, 113–140.

FREUD, S. *The interpretation of dreams*. (Originally published 1900) Vols. 4 & 5. London: Hogarth, 1953. (a)

FREUD, S. *Three essays on sexuality*. (Originally published 1905) Vol. 7. London: Hogarth, 1953. (b)

GOSSLEE, D. G. Level of significance and power of the unweighted means test. Paper read at American Statistical Association, Chicago, 1959.

GUTHEIL, E. A. *The handbook of dream analysis*. New York: Liveright, 1951.

HOSPERS, J. *Meaning and truth in the arts*. Chapel Hill: Univer. North Carolina Press, 1946.

JENSEN, A. R. The Maudsley Personality Inventory. *Acta psychol.*, 1958, **14**, 314–325.

MCNEMAR, Q. *Psychological statistics*. (Rev. ed.) New York: Wiley, 1955.

SNEDECOR, G. W. *Statistical methods*. (5th ed.) Ames, Iowa: Iowa State College Press, 1956.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

WALLACH, M. A. Art, science and representation: Toward an experimental psychology of aesthetics. *J. Aesthet. art Crit.*, 1959, **18**, 159–173.

WALLACH, M. A., & GAHM, RUTHELLEN C. Personality functions of graphic constriction and expansiveness. *J. Pers.*, 1960, **28**, 73–88.

WALLACH, M. A., & GREENBERG, CAROL. Personality functions of symbolic sexual arousal to music. *Psychol. Monogr.*, 1960, **74**(7, Whole No. 494).

(Received August 24, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 402–410

CHANGES IN INTERPERSONAL PERCEPTION AS A MEANS OF REDUCING COGNITIVE DISSONANCE[1]

KEITH E. DAVIS AND EDWARD E. JONES
Duke University

Among the problem areas to which Festinger's (1957) theory of cognitive dissonance may be applied, none has been as intensively explored in recent experiments as the area of attitude change. The paradigm of creating in an experimental group high dissonance that may be reduced by alterations of belief or feeling has been applied with reference to such attitudinal objects as toys (Brehm & Cohen, 1959b), consumer objects (Brehm, 1956), the interest value of a particular experimental task (Brehm & Cohen, 1959a; Festinger & Carlsmith, 1959), nonfavored vegetables (Brehm, 1959), membership in a discussion group (Aronson & Mills, 1959), marriage before 23 (Cohen, Terry, & Jones, 1959), and cheating (Mills, 1958). In spite of the heterogeneity of research settings, almost all of these studies have involved the creation of cognitive dissonance by inducing the subject (*S*) to engage in some behavior running counter to his private belief. It has been theoretically proposed and empirically verified that, above the minimum of pressure required to induce the *S* to perform the discrepant behavior, the greater the force to comply the *less* the resulting dissonance and the smaller the tendency to adjust one's private beliefs to support the act of compliance.

If an individual has no alternative but to behave in a fashion running counter to his beliefs, relatively little dissonance is created—it is as if he sees himself as the passive victim of fate. If he sees the possibility of behaving in a different fashion more consonant with his beliefs, and still performs in a manner at odds with his private feelings, more dissonance is created. Following this logic, there has been much explicit interest in the variable, "degree of choice." For example, Brehm and Cohen (1959b) found that children, asked to choose one of either two or four toys to keep, increased their liking of the chosen toy more when it was one of four (high choice) than when it was one of two (low choice). Festinger and Carlsmith (1959) found that the greater the monetary incentive for *S*s to argue the merits of a series of boring tasks, the less were the *S*s own attitudes toward the tasks likely to become more favorable. The greater apparent choice in the low incentive condition seemed to produce greater dissonance regarding the persuasion attempt, and this dissonance was reduced when the *S*s rated the tasks as more enjoyable than *S*s in the high incentive (low choice) condition.

The present experiment is both an extension of the above considerations into the area of person perception and an investigation of an alternative mode of dissonance reduction which has largely been ignored in experimental predictions from dissonance theory: the possibility of withdrawing or canceling out one's discrepant behavior instead of showing accommodating changes in attitude.

A recent paper by Jones and Thibaut (1958) emphasized the importance of interaction context in constructing any sensitive and comprehensive theory of person perception. Specifically emphasized in this paper was the role played by the perceiver in defining the situation and eliciting information from the stimulus person. In situations where one individual (A) has, for whatever reason, behaved toward another (B) in a manner which is discrepant with his private feelings about B, we might well expect the production of dissonance and a consequent change in A's attitude toward B. At least dissonance theory alerts us to this possibility and suggests the conditions under which such attitude change might occur. If Person A is pressured by strong incentives or imperative role demands to behave in an overly warm or overly hostile fashion toward B, we would expect little change in A's private impression of B. If, however, A is less certain about what is expected and perceives alternative ways of responding to B, the same extreme behavior would likely produce corresponding changes in

---

[1] The present research was made possible by a grant from the National Science Foundation (NSF-G8857).

attitude toward B through the mechanism of dissonance reduction. Thus, the "degree of choice" variable seems quite applicable to situations of interpersonal action and perception.

There are, however, occasions similar to the interaction depicted above where A is operating under restraints or incentives which are only temporarily effective. Thus the role requirements are not immutable and A can behave in a different fashion at a later time. An example of this possibility is the case of the practical joker who insults and derogates only to assure his victim, later on, that his remarks were all in jest. Or perhaps more aptly mentioned is the man with a gruff and hostile manner on initial contact who later reveals glimmerings of his heart of gold (and intended to all along).

These examples suggest a variable of more general relevance to dissonance theory. Without exception, the research attempts to create dissonance by inducing the *S* to act in a way that is discrepant from his beliefs, have treated the *S*'s behavior as irrevocable. *S* reads the unpleasant information, he agrees to copy the random numbers, he propagandizes the "naive" subject, he eats the unpleasant vegetable, etc. In all these cases the *S* sees little prospect of taking back or in any public way canceling out the discrepant behavior. Whether or not *S* actually goes through on his commitment to behave in a certain way, he does not anticipate any simple way of neutralizing the behavioral act. It seems logical to suggest that dissonance reduction through attitude change is dependent on the absence of such anticipations. Thus we would expect little dissonance in forced compliance situations where the unpleasant, alien, discrepant behavior can be withdrawn or neutralized.

One form which this undoubtedly takes in many real life situations is neutralization through exaggeration, parody, or dramatic misplaying (see Goffman's discussion of "realigning actions," 1959, p. 190). The child who says "thank you" to his host in response to his mother's imperious prompting may neutralize his remarks (and preserve his autonomy) through sarcastic inflection. The Korean POW who said "I am sorry I called Comrade Wong a *no-good-son-of-a-bitch*" is also effectively canceling out his compliance by emphasizing the wrong words (see Schein, 1958, p. 322).

A more direct form of cancellation, however, is the anticipation that there will be some later opportunity to take back the behavior, to say "I didn't mean it," or "My fingers were crossed," or "I was playing a joke." It is the opportunity for this latter form of neutralization which was varied in the present experiment.

To give a synopsis of the present experiment, *S*s were cajoled (Choice) or assigned (No Choice) to read aloud a negative evaluation of a stimulus person (SP) who could hear but not see the *S*. Half of the *S*s expected to meet the SP later with the experimenter (*E*), at which time *S*'s role and the deceptions involved would be explained to SP (Anticipation condition). The remaining *S*s were led to believe that such a meeting was impossible and that the SP would not be disabused (Nonanticipation condition). In line with the reasoning presented above, the main experimental prediction was: Since dissonance is greatest in that condition where *S* has the illusion of choice and is not given the opportunity to neutralize his behavior, *S* is more negative in his ratings of SP in the Choice-Nonanticipation condition than in any of the remaining three. It was difficult to make specific predictions for the other conditions except for the expectation that the least dissonance and therefore the more positive ratings would occur in the No Choice-Anticipation condition.

METHOD

Subjects

Exclusive of *S*s used in pretesting, 52 male introductory psychology student volunteers participated in the experiment.[2] The data of 12 *S*s were excluded from the analyses reported below—6 of these were suspicious that the SP was not as presented, and 6 either balked at the task presented them, or later indicated that they were subjectively in a different condition from that implied by the induction. Suspicion and subjective misclassification were evenly distributed across the four treatments. Only in the Choice-Nonanticipation condition, however, did any *S* balk at reading the evaluation—3 out of 14 *S*s run in this condition refused the induction attempt.[3] The total number of *S*s providing analyzable data was thus 40, 10 appearing in each of the four experiment conditions.

---

[2] Forty-five of these *S*s were summer school students at the University of North Carolina; the remainder were Duke summer students. We are greatly indebted to the staff of the Organization Research Group of the University of North Carolina for making their research facilities available, and particularly to Joanne Landau who was of invaluable assistance in securing *S*s.

Procedure

The procedure was administered to *S*s one at a time but *S*s were instructed that they were taking part in an experiment in first impression formation and that their task was to come to a first impression judgment of "another subject in the next room." *S*s were led to believe that *E* had taken elaborate precautions to see to it that a "genuine first impression situation would be created," and that these precautions included the pairing up of *S*s who were not taking the same courses, who were not living in the same dormitories, and who had been instructed to arrive at the experimental laboratory at slightly different times to avoid overlap in the halls, etc. Upon arrival of the *S*, *E* conducted him to an observation room behind a one-way mirror which, however, was closed off by a curtain so that the *S* could not see into the adjacent room. *E* explained that he was interested in studying only the impression created by "what a person says and how he says it, and not in the contribution of personal appearance to the impression created by a person." Then *E* asked each *S* to fill out a self-rating scale composed of 10 pairs of conventional antonyms, including generous-stingy, honest-deceitful, and timid-bold. The *S* indicated his self-rating by placing a checkmark in one of five blocks arrayed from one member of the pair to its opposite.

E then left the naive *S* alone in one room on the pretext of going to explain the experiment to the SP in an adjacent room. As soon as he reached the second room, *E* turned on a tape recording which broadcast back to *S*'s room a discussion between *E* and the SP. On the tape, *E* explained the nature of the experiment in the same terms he had used with *S*. *E* asked the SP to fill out a self-rating scale and then requested him to respond orally to a series of standard questions about his background, his family, his current interests, etc. and to three moral dilemmas adapted from the Universalism-Particularism Scale (Stouffer, 1949). SP was instructed that his answers to these questions would serve as information for *S* to use in coming to his first impression judgment of the SP, and that after SP had finished responding to these questions *S* would write down a brief personal evaluation of the SP and read this to him over the loudspeaker. As a rationale for this procedure, the SP was told that this would make the experiment "more interesting for you and allow you to hear an honest evaluation of yourself by someone who doesn't know you."

After these remarks, *E* turned off the tape, supposedly leaving the SP to organize his thoughts about the questions which he was to answer and returned to the naive *S*. While waiting for SP to begin, *E* gave *S* a carbon copy of the questions that SP was to answer and made a few informal remarks designed to clarify the procedure. After a short pause *E* returned to the SP and played the second portion of taped material consisting of SP's answers to the standard set of questions and his resolution of the representative moral dilemmas. Because the stimulus information was identical for all *S*s, and because the focus of this study was on the change scores rather than the *S*s' initial impression, no detailed description of the stimulus materials will be presented. The aim was to convey little about the SP except that he was an average, moderately likeable undergraduate. In his appearances on tape, SP seemed slightly nervous; he halted occasionally, groped for words, and seemed quite involved in creating a favorable yet honest impression of himself.

At the close of the SP's talk, *E* turned off the tape and asked the *S* to fill out a "first impression rating scale" (to be described below). Then followed the rationale for the derogation and the Choice and Anticipation manipulations:

> As you will remember, I said at the beginning of the experiment that we are also interested in how people respond to evaluations of themselves. Particularly we are interested in studying the effect on a person of extreme evaluations of himself—evaluations that are either very flattering and complimentary or negative in content. . . . What I have done is to prepare in advance two quite general evaluations, one of which is heard by every subject who participates in the other half of this experiment. One of these is quite complimentary and contains statements such as, "You sound like one of the most interesting persons that I have met since I came to Chapel Hill (Duke)," and "I would really like to get to know you much better." The other is negative and contains statements such as, "My overall impression was not too favorable," and "I wouldn't go out of my way to get to know you."

The *S*s were then randomly assigned to one of four treatments involving variations in choice and anticipation of interaction. The various sets of instructions were as follows:

> *Choice.* As far as I am concerned, you may read either of these evaluations; but I would like to ask a favor of you. Surprisingly enough, most of my recent subjects have been choosing to read the flattering evaluation of their person, so that I have not been getting much information about the effect of a negative evaluation on people. If you feel able, I'd like to ask you to read the negative evaluation of your partner. Do you think that you'll be able to do this for me?

> *No Choice.* Because we are requiring all of our

---

[3] One *S* in the No Choice-Anticipation or low dissonance condition was, in a sense, self-selected. Originally scheduled for the Choice-Nonanticipation condition, he refused to read the negative evaluation until a meeting was guaranteed (Anticipation) and before being subjected to considerable further pressure (No Choice). Thus in terms of the *final* induction and his subjective placement on the post-experimental questionnaire, this *S* fell in the No Choice-Anticipation condition. The results are quite unchanged by the inclusion or exclusion of this fortieth *S*. For ease of computation and presentation, this *S* was retained.

> *S*s to read one or the other of these evaluations, we have decided that the only fair way to assign one of these evaluations to you is to make it simply a matter of chance. Thus we have decided in advance that all even numbered subjects would be required to read the complimentary evaluation of their partner and that all odd numbered subjects would be required to read the negative evaluation of their partner. Since you are the seventh subject who has taken part in this experiment, here is the evaluation that we want you to read.

E then handed the mimeographed copy of the negative evaluation to all *S*s and began to talk about it only to interrupt himself in one of two ways:

> *Anticipation of interaction.* Of course, you realize that as soon as we have had time to study the effect of your reading of this negative evaluation on him, I shall explain to him that this was not your true opinion of him, and you will have an opportunity to meet him and to say anything that you like about your true impression of him. Of course, you don't have to say anything if you don't want to, but you will meet him as soon as we have gathered our data on his reaction.

> *Nonanticipation of interaction.* You realize, of course, that we would like to be able to tell this person that this does not represent your true opinion and that it was all part of our experimental setup, but we have found that we cannot do that and continue to do research on this topic. Therefore, we have made this kind of arrangement to protect your identity. As soon as you finish reading this evaluation to him over the microphone, I shall go into the next room and turn him over to an assistant of mine who will take him to a different part of the building where he will, by the use of interviewing and various psychological tests, get at the effect of this evaluation on him. This session will take at least thirty minutes which will give us plenty of time to finish up this part of the experiment and to see to it that you get out of the building without having to meet him and answer any embarrassing questions. Of course, he will never know your name just as you will never know his.

After *S* had had time to read the negative evaluation twice, *E* connected a microphone to an amplifier and appeared to talk briefly with the SP before turning the microphone over to the *S* for him to read it aloud to the SP. The complete negative evaluation follows below (the italicized portion was always dubbed in in pencil to make *S* feel that the evaluation varied to some extent from SP to SP):

> As I understand it, my job is to tell you in all honesty what my first impression of you is. So here goes: I hope that what I say won't cause any hard feelings but I'll have to say right away that my overall impression was not too favorable. To put it simply, I wouldn't go out of my way to get to know you. Maybe I'd change my mind if we could talk together in a more natural surrounding, but from the way you spoke—not so much what you said but how you said it—I'd guess that you have some personal problems that would make it hard for us to get along very well. Your general interests and so on just strike me as those of a pretty shallow person.

> To be more specific: *Frankly, I just wouldn't know how much I could trust you as a friend after hearing your answers to those moral questions. You took the easy way out every time.* I guess that I should point out some of the things that you said that made a good impression on me, but that would be a kind of a waste of time since the general impression that I have is not too good. That's all that I have to say.

Before leaving the room to take charge of the SP, *E* handed *S* a completed self-rating form which presumably was the one which had been filled out by the SP. *S* was asked to rerate the SP in an attempt to integrate this new information with the information already provided in the "interview." Thus, the second rating was deliberately divorced from the reading of the derogation and given a plausible justification since psychologists "often must combine information from different sources to formulate their conceptions of personality." The answers indicated on SP's self-rating scales were designed to make him appear as typical as possible. In the first place, the items themselves were chosen because it was felt that everyone would respond to them in more or less the same way. In the second place, SP's self-rating indicated that he had a positive self-picture without, however, being especially arrogant or conceited. *S*s were specifically instructed that the amount of information that they could gain from such a self-rating was limited by the willingness of the other person to reveal things about himself and that they should not feel constrained to agree with the SP about himself.

E then left the room for approximately 4 minutes after which time he returned to administer a written questionnaire designed to check on the effectiveness of the experimental manipulations and an oral questionnaire designed to get at their reaction to the SP's self-rating, to recheck the written questions about experimental manipulations, and to ascertain whether or not any *S*s were suspicious of the procedure. This oral questioning period was followed immediately by an explanation of the deception and purposes of the experiment. Most *S*s were quite relieved to learn that no one had heard the negative evaluation and left the experimental room with no apparent signs of tension.

Rating Scale

The main dependent variable was the amount of change in each *S*'s evaluation of the SP from before the derogation to after it. The rating scale was constructed specifically to reflect the *S*s' impression of the SP, and the *before* and *after* forms of the scale were identical, being composed of five trait clusters—likability, warmth, conceit, intelligence, and adjustment—each of which consisted of four simple declarative sentences, two of which were worded positively and two negatively. For example, the conceit cluster was composed of these sentences: "He strikes me as somewhat arrogant and conceited" (negative), "I think that he is basically a modest, unpretentious person" (positive), "This person seems somewhat distant and aloof" (negative), and "He seems very humble and self-effacing to me" (positive). The *S*s' ratings were made on a six-point scale reflecting the degree of agreement with each of the 20 sentences. For purposes of the present study, the ratings were algebraically summed to provide an overall measure of favorability—that is, all positive items were summed and this total was added to the inversely scored sum of scores for all negative items.

Evidence indicates that both the scale itself and scale change scores are sufficiently reliable. A split-half correlation on the before scores in which one half was composed of the first positive and the first negative item in each trait cluster yielded a .91 value, corrected by the Spearman-Brown formula. A Kuder-Richardson reliability value of .90 was also obtained on the before scores. For the change scores the reliability was somewhat lower but still quite acceptable: .60 when measured by the split-half technique and corrected by the Spearman-Brown formula, and .71 when measured by the Kuder-Richardson technique.

RESULTS

Validation of Experimental Inductions

As indicated above, only those *S*s who saw themselves in the condition assigned to them (as reflected in their responses to the post-experimental questionnaire) were retained for analysis. Without exception, all *S*s in the Anticipation conditions indicated a definite expectation of meeting the SP immediately after the experiment. None of the *S*s in the Nonanticipation condition did. Nineteen of the 20 *S*s in the No Choice condition felt that "I had no alternative but to give the evaluation I gave." The remaining *S* in No Choice and three *S*s in the Choice condition felt that they had some freedom but that there was "a great deal of pressure on me to choose the evaluation I gave." The remaining Choice *S*s scattered their responses over three alternatives reflecting greater degrees of subjective choice or freedom to choose. Since there was little or no overlap between the questionnaire responses of the retained *S*s in different conditions on questions relevant to condition differences, the experimental inductions may be considered valid almost by definition. Since only six *S*s excluded themselves by their questionnaire responses, indicating that subjectively they were not in the induced condition, the experimental inductions may be considered basically valid even without regard to *S* selection. It should also be noted, again with reference to the post-experimental questionnaire, that there was no interaction between subjective choice and anticipation. For example, the Choice *S*s in the Anticipation condition felt just as much choice as the Choice *S*s in the Nonanticipation condition.

A final matter of importance concerns the extent to which the evaluative reading was actually seen to be discrepant—in a negative direction—from each *S*'s initial attitude. Except for two *S*s in the high dissonance condition, all *S*s saw the evaluation as less favorable in tone than their own attitude. Most *S*s thought the discrepancy was considerable.

Effects of Choice and Anticipation on Rating Changes

The dependent variable of critical interest is the summary rating change score for each *S*. The means and standard deviations of these change scores are presented for each experimental condition in Table 1. A simple analysis of variance indicates that the difference between all four means is significant ($p < .05$). The variances are quite homogeneous, though there is some tendency for the *S*s in the Anticipation conditions to show greater variability than those in the Nonanticipation conditions.

As always, interpretation of significant differences in change scores depends on the relationship between change scores and before scores. In the present case, the differences between the four before-score means did not approach significance. Also, none of the four

TABLE 1

MEAN BEFORE AND CHANGE SCORES AND STANDARD DEVIATIONS OF SUMMARY RATING EVALUATIONS

	Condition			
	Choice		No Choice	
	Antic	Nonantic	Antic	Nonantic
Before[a]				
M	51.0	55.5	48.0	54.7
SD	8.92	17.31	11.89	15.25
Change[b]				
M	−1.8	−7.7[c]	−2.2	−1.7
SD	6.63	4.79	5.39	3.43

Note.—The lower the before score the more favorable the rating. A minus change score indicates change toward a less favorable impression. $N = 10$ for each cell; total $N = 40$.

[a] F betw/with = .641.

[b] F betw/with = 3.41,/3 and 36 *df*, $p < .05$.

[c] Individual mean comparisons with C-Na (Choice Nonanticipation)—NC-Na: $t = 3.222$, 19 *df*, $p < .01$; C-A: $t = 2.278$, 19 *df*, $p < .05$; NC-A: $t = 2.413$, 19 *df*, $p < .05$.

correlations between before scores and change scores differed significantly from zero; two were in the positive, two were in the negative direction. As a final check on the possibility that the change scores were influenced by the before scores in some systematic way differing across treatments, an analysis of covariance was done which indicated that the difference between means was still significant after partialing out any effects of before score.

The role of choice. The main hypothesis of the experiment was that the illusion of choosing to read the negative evaluation plus the awareness that there would be no disabusing interaction with the SP would combine to produce the greatest amount of dissonance. The most direct consequence of this heightened dissonance would be a high negative change on the summary evaluation score, relative to the other three experimental groups. It is clear by inspection of Table 1 that this main hypothesis is confirmed. The Choice-Nonanticipation group mean is significantly different from each of the other group means at the 5% level or lower. The remaining three groups are obviously very similar in mean and the differences between them do not approach significance.

It is somewhat surprising to note that the No Choice-Anticipation condition, expected to produce the least dissonance, results in as much negative change as do the two moderate dissonance conditions (Choice-Anticipation and No Choice-Nonanticipation). Apparently, whether or not one has any freedom to read the negative evaluation is not related to the production of dissonance as long as *S* feels he can retract his behavior after the experiment. When this avenue of dissonance reduction (or prevention) is closed, however, the present experiment clearly confirms the role of choice or commitment in the production of dissonance.

The role of anticipation. In view of the pattern of rating change results, one could make the same sort of remarks about the role of anticipation as about the role of choice: knowledge that there will be no interaction is essential in the production of dissonance only when *S* has the illusion of some choice in reading the evaluation. However, as we might infer by the tendency toward greater variability in the anticipation groups, the anticipation variable is probably more complex than the choice variable—or at least the lines between conceptual status and operational definition are more tenuous with regard to anticipation than choice. Choosing to behave in a certain manner is a rather direct index of committing oneself to the behavior, of "owning up to it," or acknowledging it is at least partially self-caused.

In the present experiment, the anticipation of meeting or not meeting the SP was manipulated to allow half of the *S*s to reduce or prevent dissonance by anticipating the withdrawal of their behavior rather than changing their attitude. But anticipation had other meanings to these *S*s as well. A few *S*s openly expressed anxiety at the prospect of meeting the SP. As one *S* not too lightly said: "I hope he's not a big guy." There were probably others who shared some of these feelings but did not express them openly. Nevertheless, there is evidence from the post-experimental questionnaire that such anxiety reactions were definitely in the minority. When asked, "How appealing is (or would be) the prospect of meeting the other subject to explain your part in this?" 10 (or half) of the Anticipation *S*s checked the alternative, "very much like to meet him and explain." Six more *S*s checked the next alternative, "rather like to meet him and explain." Only three *S*s checked the third or neutral alternative, with one saying "rather not meet him . . . ." In contrast to this, *S*s in the Nonanticipation condition tended to feel much more neutral about meeting the SP: 11 out of 20 checking the neutral alternative. The *S*s in the Anticipation conditions are, in fact, significantly *more* eager to meet the SP than the *S*s in the Nonanticipation conditions ($\chi^2 = 4.401$, 1 *df*, corrected for continuity, $p < .05$). This suggests, perhaps, a "sour grapes" reaction in the nonanticipation *S*s—since they cannot meet the SP anyway, the dissonance produced by this is partly reduced by expressing a lack of interest in such a meeting. In any event, the relative eagerness of the anticipation *S*s to meet the SP indicates that anticipation of interaction plays a greater part in dissonance reduction than in anxiety arousal, and that the latter reaction is atypical.

Performance Variations and Attributions of Suspicion

It was impossible to control the facility and sincerity with which the *S*s read the derogatory evaluation, except by strong instructions urging simulated sincerity and the provision of a brief practice reading period. It is conceivable that some *S*s deliberately tried to sound insincere or to give cues to the SP that their performance was other than genuine. This possibility was mentioned in the introduction as an alternative means of discounting behavior which is discrepant from other established cognitions. For this reason, *E* attempted to rate each performer on the sincerity of his performance. While it turned out to be impossible to make reliable ratings of such an intangible quality, *E* was aware of only one *S* who seemed to be deliberately conveying cues by exaggerating his part over the microphone. Probably because of the constraints inherent in the situation, then, attempts at caricature in the reading performance did not serve as a major means of dissonance reduction in the present experiment.

Nevertheless, *S*s did vary in response to one question which asked whether they thought SP was likely to be suspicious of the sincerity or source of the evaluation which was read to them. Fifteen of the *S*s felt that the SP was likely to be suspicious, and these distributed themselves quite evenly across the different conditions. Since the attribution of suspicion is a possible way of reducing dissonance, it is important to note the relationship between such attribution and rating change scores. Here a rather curious and unexpected finding emerges. In the Nonanticipation conditions, there is no relationship between attribution of suspicion and change score. In the Anticipation conditions, those who attribute suspicion to the SP show significantly greater negative rating change than those who do not ($t = 3.41$, $p < .01$). In none of the comparisons does the choice variable show any relationship to the attribution of suspicion.

The difference within the Anticipation conditions seems paradoxical and quite difficult to explain. One would assume that attributing suspicion to SP would be one way of neutralizing the situation and avoiding dissonance; yet those who attribute suspicion are precisely the ones who show the greatest dissonance reduction by changes in rating. There are a number of equally plausible explanations for this finding which are neither refuted nor supported by available data. Any commitment to a particular explanation would therefore be quite arbitrary. In short, the findings with regard to attribution of suspicion remain essentially a mystery.

The Role of Self-Rating Discrepancy

In order to support the rationale of the experiment, it was essential to provide some further information about the SP during the interval between the first and second rating of him. An attempt was made to construct a bogus self-rating performance for the SP which was purposely uninformative and yet could serve as a minimal justification for changes in rating. Some information was provided by this SP self-rating, however, and many *S*s claimed that they used this information in re-evaluating him. Many of those who became more negative in their ratings felt that the SP rated himself in a conceited or egocentric way. Since the attribution of conceit would presumably bear some relation to one's own self-ratings on the same scale, the discrepancy between the *S*'s and the bogus SP's ratings was considered as a possibly confounding variable. If, for example, *S*s in the Choice-Nonanticipation condition were for some reason more modest in their self-ratings than *S*s in the other conditions, they would tend to think of the SP as relatively conceited and therefore dislike him. Such a chance occurrence would make it unnecessary to invoke dissonance theory in order to explain the results.

There were variations in average self-rating discrepancy from group to group, and it so happens that the largest discrepancy occurred in the critical Choice-Nonanticipation condition. None of these differences in self-rating discrepancy means departed significantly from chance, but as a final control over this factor covariance analyses were conducted to partial out the effects of self-rating discrepancy in comparing Choice-Nonanticipation with each other condition in turn.

Suffice it to say that there were no essential changes in results or significance levels except for the overall between groups effect, which drops just below the 5% level. Since this overall comparison includes some groups between which no differences were predicted, and the crucial *t* test comparisons (Choice-Nonanticipation versus the other groups) remain essentially unchanged, we may conclude that the present results cannot be explained by chance variation in the self-ratings of *S*s assigned to different conditions.

DISCUSSION

The results of the present experiment indicate the conditions under which a person, whose behavior toward another departs from his initial attitude toward the other, changes his attitude to make it consonant with his behavior. The person or actor must (*a*) feel that he had some freedom *not* to behave in the discrepant manner, and (*b*) realize that he cannot easily disclaim the behavior in the eyes of the target person. According to the findings presented above, the combination of these two conditions is necessary and sufficient to produce enough dissonance to motivate attitude change. In the absence of either condition, the amount of attitude change is negligible and significantly less than when both are present. We assume in this general statement that it is merely the amount of discrepancy between behavior and initial attitude which is important and not the direction of discrepancy. Since the present experiment only involved behavior which was more negative than the initial attitude, the generality of the present results is contingent on further empirical investigation.

The first of the two necessary conditions (Choice) has received considerable attention in the research of others. The present experiment thus confirms and extends to a different attitudinal object (a person) the findings of Brehm and Cohen (1959a, 1959b) and of Festinger and Carlsmith (1959). The present experiment indicates, however, that the choice to behave in a certain manner must have fairly irrevocable consequences in order for dissonance and consequent attitude change to appear. When *S*s are led to believe that they will soon meet the target of their negative evaluation, and that they may then explain away their behavior, the choice variable has no effect on attitude change. In considering the consequence of "forced compliance," therefore, it is essential to note the available means of neutralizing the behavioral act involved in the expected dissonance. Can the person easily change the meaning of this act or modify its significance? Can he later disown or revoke the behavior? One effect of such possibilities, specifically the latter one, is brought to light by the present results. The commitment implicit in choosing to behave in a manner discrepant from attitude, must be a commitment to a durable, preferably public performance, and not a series of acts which can easily be neutralized or taken back.

The implications of the present results for theories of social interaction and perception are rather straightforward. Certainly some clues are provided for predicting the effects of interpersonal communication on interpersonal attitude. A hostile comment that is clearly demanded by role requirements (No Choice) should lead to less change in the communicator's attitude than a hostile comment that represents a response to more ambiguous circumstances (Choice). From this we might further infer that the well practiced performer, the man who feels every interpersonal act is "spelled out in the manual," and who knows the pattern of behavior required by each role he is forced to play—such a man would not likely be affected by his own utterances toward others. On the other hand, the improviser, or the person who is uncertain of situational and role requirements, is much more apt to be constantly in the throes of dissonance elevation and reduction. In the process, presumably, his impressions of the same other may actually be quite variable in the course of an extended conversation. The potential dilemma of this latter type of person is blunted by the fact that he may succeed in always behaving toward another in a manner consonant with his initial attitude. Given some (preferably ambiguous) pressure to depart from this consonance of communicative behavior, however, there should be clear fluctuations in attitude or personal impression.

SUMMARY

This study was designed to demonstrate that dissonance produced by inconsistencies between one's evaluation of another and one's behavior toward him would be reduced, under certain conditions only, by changing one's evaluation to conform to the behavior. All *S*s read a negative evaluation about a stimulus person over a microphone presumably broadcasting to him. *S*s were instructed to act as though this was their true evaluation of the stimulus person. Half of the *S*s (Choice conditions) were given some opportunity to read a positive evaluation but urged to read the negative one; the remaining *S*s (No Choice conditions) were simply assigned to read the negative evaluation. Half of the *S*s in each of these groups were in turn informed that they would have a chance to meet the stimulus person immediately after the experiment when all would be explained to him (Anticipation conditions); the remaining *S*s were told that no meeting would be possible and that the stimulus person could not be informed of the deception (Nonanticipation conditions).

The major prediction derived from dissonance theory was that only those *S*s in the Choice-Nonanticipation condition would have sufficient dissonance so that a negative change in the impressional evaluation of the stimulus person would occur. This prediction was confirmed, though it was surprising that the No Choice-Anticipation group showed no less attitude change than the "moderate dissonance" conditions: No Choice-Nonanticipation and Choice-Anticipation. Some incidental findings were presented and the results were discussed in terms of dissonance theory and implications for interpersonal perception.

REFERENCES

ARONSON, E., & MILLS, J. The effect of severity of initiation on liking for a group. *J. abnorm. soc. Psychol.*, 1959, **59**, 177–182.

BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, **52**, 384–389.

BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379–382.

BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, **58**, 383–387. (a)

BREHM, J. W., & COHEN, A. R. Re-evaluation of choice alternatives as a function of their number and qualitative similarity. *J. abnorm. soc. Psychol.*, 1959, **58**, 373-378. (b)

COHEN, A. R., TERRY, H. I., & JONES, C. B. Attitudinal effects of choice in exposure to counterpropaganda. *J. abnorm. soc. Psychol.*, 1959, **58**, 388–391.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203–210.

GOFFMAN, E. *The presentation of self in everyday life.* Garden City, N. Y.: Doubleday Anchor, 1959.

JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958. Pp. 151–179.

MILLS, J. Changes in moral attitudes following temptation. *J. Pers.*, 1958, **26**, 517–531.

SCHEIN, E. H. The Chinese indoctrination program for prisoners of war: A study of attempted "brainwashing." In Eleanor Maccoby, T. Newcomb, & E. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1958. Pp. 311–344.

STOUFFER, S. A. An analysis of conflicting social norms. *Amer. sociol. Rev.*, 1949, **14**, 707–717.

(Received August 31, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 411-418

SOME PERSONAL AND SITUATIONAL DETERMINANTS OF YIELDING TO INFLUENCE[1]

AMBALAL S. PATEL[2] AND JESSE E. GORDON
University of Wisconsin *University of Michigan*

THE purpose of this study was to discover some of the personal and social variables that are related to yielding to influence in a situation in which subjects (*S*s) are given the option of yielding to or ignoring the influence.

In typical experiments in this area, *S*s are asked to make judgments before and after experimentally produced "false" judgments are presented; changes in *S*s judgments in the direction of the artificial ones are taken as evidence for suggestibility or yielding to influence (Moore, 1921; Sherif, 1936). Studies in which judgments provide the dependent measures, such as literary judgments, judgments of line lengths or of extent of movement in the autokinetic effect, yield evidence that many *S*s modify their verbal behavior to conform to group norms or to the prestige suggestions. However, such studies provide little evidence that *S*s do in fact change their attitudes sufficiently for there to be a change in their overt behavior outside of the experimental situation. It is possible that *S*s, usually both curious and uncomfortable in experiments in which they participate, were in the position of searching not for influence but for guidance in responding appropriately. The experimenter produced cues, which are usually labeled as prestige suggestions or social influence, may be perceived by *S*s as guides or cues for structuring the task and thus reducing *S*s' trial and error. Operating on such an assumption, *S*s may give verbal responses of a conforming nature in order to avoid censure, without there being any more generalized effects of the influence on their judgments or response tendencies outside of the situation. When Hamlet leads Polonius to agree that a cloud has in turn the shape of a camel, a weasel, and a whale, we understand that he is extracting a response of expediency and not necessarily producing a change in Polonius' perceptions of clouds.

To an extent this problem is related to the deliberate subtlety with which the influence is introduced in the experiments. The situations are kept vague so that *S*s will not discover the purpose of the experiment and will not know that they are being subjected to influence. This vagueness is made necessary by the experimenter's (*E*'s) fears that should *S* know that he is being subjected to influence, he would resist it. Such experiments do throw considerable light on the extent to which a person's verbal behavior may be controlled by the cues present in a situation, but it cannot be said that *S*s are yielding to influence or accepting suggestions when, as far as *S* is concerned, no suggestions have been offered.

An experiment by Coffin (1941) corrects this situation to some extent, by making the influence explicit. He asked his *S*s to solve mathematical problems, and penciled in both correct and incorrect "hints" for the solutions of the problems. However, even here *S*s were not free to refuse the suggested hints, because Coffin announced to his *S*s that he had given hints in order to make the problems easier to solve. Assuming the rapport that exists between students and a teacher-researcher, it would seem that Coffin's *S*s could better be described as following instructions, both explicit and implicit, when they used the "hints" given them, than as yielding to influence. It is therefore not surprising that those *S*s with little knowledge of mathematics employed more hints: they did not have the knowledge necessary to recognize that the hints were misleading in some cases, and considering their relationship with the teacher-experimenter, assumed that the hints were valid guides to the solutions of the problems.

---

[1] This paper was based on a thesis which was submitted in partial fulfillment of the requirements for the MS Degree in the Department of Education of the University of Wisconsin.

The authors wish to extend their appreciation to A. H. Edgerton of the Department of Education and D. A. Grant of the Department of Psychology for their assistance in this study, and to the Wisconsin Alumni Research Foundation for its financial aid in support of this research.

[2] Now at the University of Baroda, Baroda, India.

This problem is a part of the relative importance of normative and informational influence. These terms, proposed by Deutsch and Gerard (1955), identify different aspects of suggestions: a suggestion is informational when it transmits some presumable trustworthy data about the to-be-judged stimuli to an *S*; a suggestion is normative when it represents a response that is characteristic of a group. It is obviously quite difficult to separate these aspects from each other. A suggestion that is informational usually comes from a source, thus bringing with it the implication that it carries the authority associated with the prestige of the source person and the group to which he belongs, and is thus normative. In general, the informational value of a suggestion varies with *S*s' state of knowledge; the normative value varies with the relationship between *S* and the source of the suggestion. Deutsch and Gerard (1955) attempted to separate these aspects of suggestion by having some *S*s report their judgments anonymously, after having been subjected to influence, while others made their judgments with varying degrees of public commitment to the influencing group. They found some modification of judgment even in the anonymous group, in which normative influence was presumably not operating, and suggest that this modification is a result of informational influence. However, there was no effort to manipulate the state of *S*'s own information regarding the stimulus; without such manipulation, it is not possible definitely to identify the influence as informational in the anonymous group. It is possible that even though *S* was not publicly committed to his judgment, his desire to be like the "stooges" may have resulted in the modification of his judgments. If this were the case, then the suggestions should be considered normative rather than informational. Samelson (1957) tested and confirmed the hypothesis, derived from this distinction between normative and informational influence, that where *S* was given reason to doubt the accuracy of the source (thus reducing its informational value), conforming behavior would be reduced. These two studies thus provide evidence that both informational and normative suggestions are capable of modifying judgment, although Deutsch and Gerard were not able to manipulate normative suggestions without informational components being included.

The study reported here throws some light on the relationships among informational and normative influences and acceptance of influence in a situation in which the interactions of these influences may be observed. The informational value of influence is manipulated by varying the initial level of information of the *S*s relative to the task; the normative value of the same influence is manipulated by ascribing the influence to high and low prestige sources. Under these conditions, if it is hypothesized that both aspects of influence produce modifications in *S*s' behavior, then high prestige influence directed at the low knowledgeable *S*s should produce high acceptance of the influence, and low prestige influence directed at the high knowledgeable *S*s should produce little acceptance of the influence. *S*s are made explicitly aware that suggestions are present, and that they may choose to accept or reject them. The procedures were designed to eliminate the implication that the suggestions presented were necessarily informational, rather than normative, and that they constitute a valid guide for behavior in the experiment.

The experimental task facing the *S*s in this research was to indicate the correct synonyms for English words, the stimulus words being graded in difficulty. Because the test was administered in high school English classes by the English teachers,[3] and was not introduced as an experiment, it seems likely that *S*s' responses to the test items, whether or not they accepted the suggestions given, represented their honest belief that the words they checked as synonymous were in fact so. It is doubtful that their choices were influenced by the kind of expediency illustrated by Polonius' responses to Hamlet, and which may have been operating in "group pressure" experiments. It is more likely, therefore, that *S*s' acceptance of suggestion or influence produced changes not only in their verbal responses, but also in their cognitive functioning related to the influenced responses.

The independent variables whose relation-

---

[3] The authors wish to acknowledge the assistance and cooperation of the administrators and teachers of the Portage High School, Portage, Wisconsin.

TABLE 1
ANALYSIS OF VARIANCE OF SUGGESTIBILITY SCORES

Source	SS	df	MS	F
Main Effects:				
Grade	13837.10	2	6918.55	17.28***
Prestige	1928.01	1	1928.01	4.81*
Sex	2821.47	1	2821.47	7.05**
Difficulty	14890.30	1	14890.30	37.18***
Interactions:				
Grade × Prestige	3659.40	2	1829.70	4.57*
Grade × Prestige × Sex	3614.02	2	1807.01	4.51*
Grade × Prestige × Difficulty	4405.39	2	2202.69	5.50**
Grade × Prestige × Sex × Difficulty	4752.92	2	2376.46	5.03**
Error:				
Within Groups	48054.69	120	400.46	
Total:	98139.78	143		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

ships to the acceptance of suggestions were measured were: prestige of the source of influence, ambiguity of the problems for which suggested solutions were provided, sex of *S*, and age-grade of *S*, varied independently of ambiguity of the problems.

METHOD

Ss were 72 high school students, from the tenth, eleventh, and twelfth grades. Both sexes were represented in each grade.

Ss were given Lepley's Synonym Vocabulary Test (Lepley, 1955),[4] consisting of eight parts. Each part contains 30 stimulus words and 5 response words from which *S* must pick the correct one for each stimulus word by marking A, B, C, D, or E on an IBM answer sheet. Correctness of response is in terms of similarity in meaning to the stimulus word. The only differences among the eight parts are in the stimulus and response words, and their difficulty as determined from Lepley's standardization groups.

Suggestions were given *Ss* by marking in pencil in the questionnaire booklets an incorrect letter next to 50 of the stimulus words, and a correct letter next to 11 of the stimulus words. The first 4 stimulus words marked were marked correctly, in order to create confidence in the markings by the *Ss*. Thereafter, in the test, the remaining 7 correct marks were randomly distributed among the 50 incorrect ones. The markings were identical for all *Ss*. Only the incorrectly suggested items were used for the analysis of the effects of suggestion.

Use of the marked "hints" was made optional for the *Ss* by telling them that they were to put their answers on the IBM answer sheets and make no marks on the questionnaire booklet. The teacher who administered the tests then indicated that the test had been taken by other classes previously, and she noted that some of the students in the other classes had put marks on the test booklets, but that these marks should be ignored. Thus the suggestions were directly provided to the *Ss*, who were explicitly aware of the suggestions, at the same time as they were permitted to accept or reject the suggestions without being subjected to pressure to accept from the supposed originator of the suggestion.

Prestige of suggestion source was manipulated by giving differing instructions to half of each grade group. In one half of each group, the teacher indicated that the class which had taken the test previously was one year lower in school than the *S* group; in the other half, it was indicated that the class which had taken the test was one year ahead of the *S* group. Thus half of the tenth grade sample was told that the test was taken just previously by ninth grade students (who were, by implication, the ones who made the marks on the test booklets), and half were told that it was taken by the eleventh grade.

Ambiguity of the problem facing the *Ss* was manipulated independently of grade level by having each *S* rate the words whose synonyms had been "suggested" several days after the test was given. Each *S* was given a list of the 61 marked stimulus words and two others as an example, and was asked to place a cross next to those which he considered to be difficult to understand in general, or whose meaning he did not know, and a check next to those whose meaning *S* knew. *S*'s scores for acceptance of influence were computed individually as the proportion of his "easy" and of his "difficult" words on which he took the incorrect suggestions.

Analysis of the data was in a 3 × 2 × 2 × 2 factorial design, consisting of three levels of age-grade status, two sexes, two levels of difficulty, and two levels of prestige source.

RESULTS

The results of the main analysis are presented in Table 1. Table 2 presents the means

---

[4] Thanks are due to W. M. Lepley for making a copy of his test available for this research.

TABLE 2

MEAN PROPORTIONS OF SUGGESTIONS ACCEPTED

Groups	Easy	Difficult	Total
1. Females:			
A. Low Prestige: Grade 10	37.95	59.12	48.54
Grade 11	24.20	46.25	35.23
Grade 12	18.95	33.77	26.36
Total	27.03	46.38	36.71
B. High Prestige: Grade 10	34.78	63.88	49.33
Grade 11	47.15	72.45	59.80
Grade 12	20.27	36.08	28.18
Total	34.07	57.47	45.77
Total All Females	30.55	51.92	41.24
2. Males:			
A. Low Prestige: Grade 10	21.20	50.16	35.68
Grade 11	22.18	39.00	30.59
Grade 12	15.63	37.69	26.66
Total	19.67	42.28	30.98
B. High Prestige: Grade 10	33.26	55.97	44.62
Grade 11	39.49	57.87	48.68
Grade 12	10.83	19.77	15.30
Total	27.86	44.54	36.20
Total All Males	23.76	43.41	33.59
3. Grand Total	27.15	47.67	37.41

of *S*s' suggestibility scores (proportion of easy and difficult items answered in the way indicated by the incorrect suggestions).

Suggestibility and Age-Grade Level. Tables 1 and 2 indicate a significant difference in the acceptance of suggestions as a function of age-grade level. A Tukey gap test applied to these data indicate that Grade 12 *S*s are significantly different from Grades 10 and 11 ($p < .001$), but these latter two grades do not differ significantly from each other. Thus the significant variance associated with Grade appears to be mainly the result of the relatively low suggestibility of Grade 12.

Suggestibility and Prestige Level. Table 1 indicates a significant effect associated with the prestige status of the suggestion source, although confidence in this effect must be lower than it is in the other tabled relationships. In general, there was a tendency for *S*s to accept suggestions more frequently if the source of the suggestion had high prestige than if the source had low prestige. This was particularly true for the Grade 11 *S*s, who were largely responsible for the significant Grade × Prestige interaction. While tenth graders accepted very few more suggestions from a high prestige source (about 5% more), and twelfth graders actually were slightly less suggestible when the prestige source was high (they accepted about 7% fewer high prestige than low prestige suggestions), the eleventh grade students were markedly affected by the high prestige, accepting about 20% more high prestige suggestions than low ones. Apparently the prestige of a high school senior has much greater influence on an eleventh grade student than does an eleventh grader on a tenth grader, or a college freshman on a high school senior. These data indicate that where prestige is low, there is a decreasing tendency to accept suggestions as one increases in age-grade status, but that high prestige can alter this trend. This strong interaction of Grade and Prestige masked the differences in suggestibility between Grades 10 and 11; when one applies the Tukey gap test to the suggestibility scores of tenth, eleventh, and twelfth grades in the low prestige condition only, one finds that each grade is significantly different from the others, in the direction of decreasing suggestibility with increasing grade ($p < .001$ in each comparison).

Suggestibility and Sex. Tables 1 and 2 indicate that girls accepted significantly more suggestions than did boys. Examination of the Sex × Prestige × Grade interaction reveals that the influence of sex was not constant over all conditions. The difference between boys and girls in suggestibility diminishes with increasing grade level in the low prestige groups, such that the girls accepted significantly more low prestige suggestions than did boys at the tenth grade, with the difference between boys and

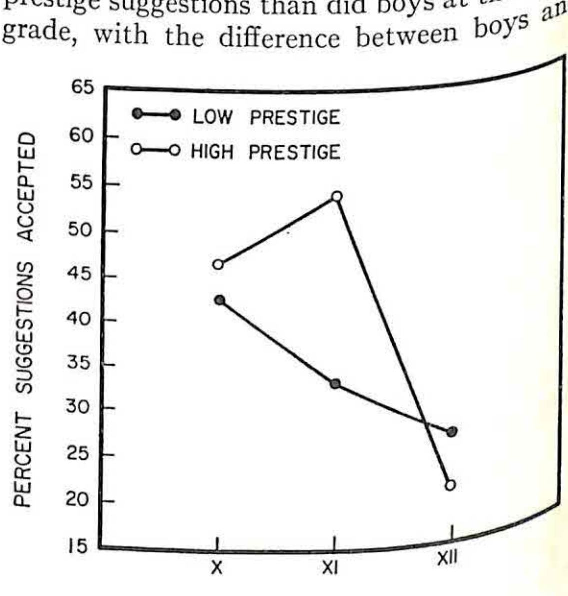

FIG. 1. Interaction of prestige and grade.

girls diminished but still significant at eleventh grade, and the difference between the sexes nonsignificant at twelfth grade. However, in the high prestige groups, girls remain more suggestible than boys throughout the three grades.

Suggestibility and Difficulty Level. Tables 1 and 2 reveal a significant difference in the direction of accepting more suggestions on items that *S*s considered to be difficult than on those they thought to be easy. Thus ambiguity of the situations in which *S* must perform is a potent factor in producing dependence on suggestions. Examination of the significant Difficulty × Prestige × Grade interaction indicates that high prestige had a greater effect on the acceptance of suggestions on difficult items in the twelfth grade *S*s than it did on easy items for that grade, whereas in the other two grades, there was no apparent interaction of prestige and difficulty of the item. In general, twelfth graders seemed little influenced by suggestions from either high or low prestige sources, when the items were considered easy. But on difficult items, they did respond differentially to the two prestige sources, so that on difficult items they accepted as many high prestige suggestions as did tenth graders. Thus difficulty and prestige in interaction were able to mask completely the general tendency of twelfth graders to be relatively independent of suggestion.

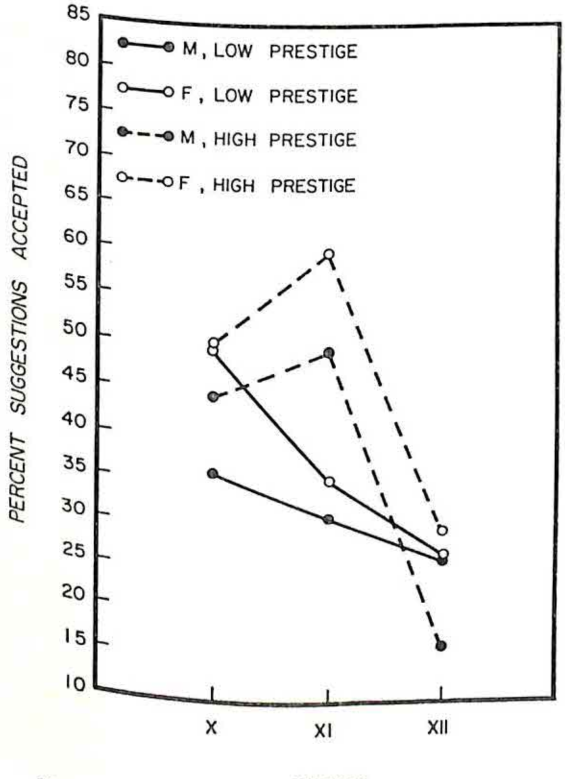

FIG. 2. Interaction of prestige, grade, and sex.

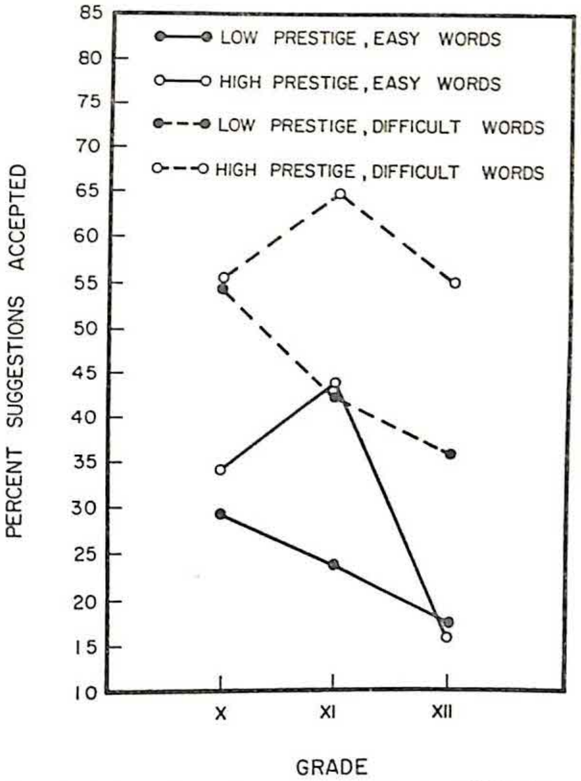

FIG. 3. Interaction of prestige, grade, and difficulty.

DISCUSSION

In this study, a grand total of 37% of the items on which misleading suggestions were provided were answered in the suggested direction. On roughly two-thirds of the items, even with suggestions provided, *S*s responded without recourse to the "hints." However, the study indicates that this over-all figure of 37% could be inflated or deflated considerably by manipulation of the proper variables, so that among tenth and eleventh grade girls given high prestige suggestions on difficult material, an average of 68% of the suggestions were accepted, and among twelfth grade boys, an average of only 23% of the suggestions made on items considered to be easy were accepted. It thus appears that sex, age-grade status, prestige status of the suggestion source, and difficulty or ambiguity of the problems facing *S*s are powerful variables in affecting *S*'s decision to accept or not accept suggested solutions.

The results of this study are in accord with the commonly-held view that at lower age-grade levels, children are more impressionable, in the sense that they are more accepting of suggestions as guides for their behavior, and that as they grow older there is an increasing tendency for them to become more independent of suggestions from others. It is of interest that this greater suggestibility at lower age-grade levels is particularly true for low prestige suggestions, implying that when young children accept suggestions, it is without reference to or evaluation of the adequacy of the source of the suggestions. As they become less suggestible, they also become more discriminating, so that they come to accept suggestions only in restricted situations: those involving difficult material and where the prestige of the suggestion source is high. Perhaps this is another way of saying that they become more self-reliant and independent. It is of interest that this trend is not based on increasing ease in solving problems, with a concurrent lowering of the need to accept suggestions. In terms of the design of this study, difficulty level of the problems was evaluated independently of age-grade status.

Eleventh graders in this study showed a somewhat peculiar pattern in their responsiveness to suggestions. When the suggestions were ascribed to a low prestige source, the eleventh grade *S*s went along with the general downward trend of suggestibility with increasing age-grade level. However, when the suggestions were ascribed to high school students in their senior year, eleventh graders accepted more suggestions than did any other group. The finding suggests that for eleventh graders, seniors may be much more prestigeful than eleventh graders are for tenth grade students or college freshmen for seniors. This is another way of saying that one year increments in age-grade status are not psychophysically equivalent or equal intervals so far as prestige effects are concerned. It is quite reasonable to suppose that in terms of the social structure of a high school, seniors occupy the highest prestige position, whereas discriminations between sophomores and juniors are not so readily made. As far as the tenth year students are concerned, juniors are not so very different from themselves, whereas the senior year is an outstanding goal for all of the students, including juniors, who are well aware of the differences between themselves and those who are about to be graduated. Further, because seniors do occupy the high prestige position in school, the effects of the status of college freshmen on themselves may be nullified by their own high self-evaluations. The lack of prestige influence by college freshmen may in part also be attributed to the fact that freshmen are perceived even by their educational inferiors as occupying a particularly nonprestigeful position in the college world, in which they are required to wear marks of their low status ("beanies") and are the continual butts of jokes and leg pulling by the more advanced students. These findings suggest that the invocation of "getting into college" as a motivating goal state by high school teachers may be less effective than referring to "being promoted to the senior year."

The increase in acceptance of suggestions as a function of difficulty level of the problems presented to *S*s provides some confirmation of the idea that conformity experiments, in which ambiguity is purposely maintained, inherently increase the probability that *S*s will accept the influence provided by "stooge" *S*s. This finding is in direct agreement with Coffin, who presented his *S*s with mathematical problems after having first categorized his *S*s in terms of amount of mathematical training, and found greater acceptance of erroneous "hints" among the poorly trained than among the well trained.

Girls in this study were in general more suggestible than boys, confirming a common cultural stereotype regarding the responsiveness of females to suggestions. The fact that the girls, as they increased in age-grade level, approached the boys in resistance to suggestions where the prestige was low, but remained consistently more responsive to high prestige than boys is also in line with this stereotype. These findings add to those of McClelland, Atkinson, Clark, and Lowell (1953) that achievement motivation is more readily aroused in girls than in boys by invoking a standard of social acceptability with which to compete, and is more readily aroused this way than by invoking a "standard of excellence." It may be that in general, American girls are more responsive to social stimulation and status than are boys. However, the greater suggestibility of girls than of boys is not the result of a failure on the part of girls to reduce their responsiveness to high prestige suggestions in twelfth grade. Such a reduction is in evidence, so that twelfth grade girls are only slightly more responsive to high prestige than to low prestige suggestions. However, twelfth grade boys are actually less responsive to high than to low prestige suggestions. The data indicate what may almost be a reaction formation against prestige in the senior boys; they accept more low prestige suggestions than they do high. This reversal, not occurring in the girls, thus makes boys less responsive to high prestige suggestions than girls, although the sexes are equally responsive to low prestige suggestions.

In general these data indicate that both normative and informational aspects of influence are capable of producing modifications in *S*s' behavior. Where *S*'s level of knowledge is low, acceptance of suggestions from both high and low prestige sources is high, and acceptance of suggestions is higher when they come from more prestigeful sources. The data also suggest that although males and females of the same grade level, working on material of equal difficulty, may be considered to have approximately the same amount of knowledge relative to the task, they differ in their acceptance of suggestions. These findings therefore imply that they differ in their responses to the normative aspects of influence, rather than the informational, such that females are more responsive to high prestige than males.

SUMMARY

An experiment was conducted using boys and girls in Grades 10, 11, and 12 to explore factors associated with the acceptance of suggestions in a situation in which *S*s were given the option of accepting or not accepting the suggestion, and in which task difficulty was varied independently of age-grade status. The variables explored were age-grade status, sex of *S*, difficulty level (or ambiguity) of the task on which *S* was to perform, and prestige level of the suggestion source. Suggestions of high and low prestige were given by penciling in the incorrect answers to some of the items on the Lepley Synonym Vocabulary Test, and ascribing the marked answers to a failure to follow directions to not mark the questions in a class that supposedly had taken the test the previous hour. Prestige level of the suggestions were varied by describing the class which had supposedly taken the test earlier as one year higher or lower in school than the *S*s. Difficulty level was evaluated by having *S*s later rate the words, for which suggested incorrect solutions were provided, as difficult or easy; each *S*'s scores were expressed as proportions of easy and difficult words on which he or she accepted the incorrect suggestions. Results were as follows:

1. Acceptance of suggestions was higher in girls than in boys.
2. High prestige suggestions were accepted more frequently than were those ascribed to a low prestige source.
3. Suggestions were accepted more frequently on difficult than on easy items.
4. Twelfth grade *S*s accepted significantly fewer suggestions than did tenth or eleventh grade *S*s, who did not differ significantly from each other, except when high prestige groups were excluded from analysis. When suggestion source was of low prestige, the three grades differed significantly from each other, in the direction of decreasing acceptance of suggestions.
5. Eleventh grade *S*s accepted many more high prestige suggestions than did tenth or twelfth grade *S*s, suggesting that seniors are more prestigeful to eleventh graders than eleventh graders are to tenth graders, or college freshmen to seniors.
6. The differences between boys and girls in acceptance of suggestions diminished with increasing age-grade status, where the prestige level of the suggestion source was low, but girls remained higher than boys in the acceptance of high prestige suggestions. Boys in the senior year were more resistant to high than to low prestige suggestions, while girls remained more responsive to high prestige.
7. In the twelfth grade, high prestige suggestions had a greater effect on difficult items than it did on easy items. Resistance to high

prestige suggestions could be completely overcome in this group by high item difficulty.

REFERENCES

COFFIN, T. E. Some conditions of suggestion and suggestibility: A study of certain attitudinal and situational factors influencing the process of suggestion. *Psychol. Monogr.*, 1941, **53**(4, Whole No. 241).

DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 620–636.

LEPLEY, W. M. The rationale, construction, and preliminary try-out of the Synonym Vocabulary Test. *J. Psychol.* 1955, **39**, 215–225.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MOORE, H. T. The comparative influence of majority and expert opinion. *Amer. J. Psychol.*, 1921, **32**, 16–20.

SAMELSON, F. Conforming behavior under two conditions of conflict in the cognitive field. *J. abnorm. soc. Psychol.*, 1957, **55**, 181–187.

SHERIF, M. *The psychology of social norms.* New York: Harper's, 1936.

(Received September 8, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 419-423

THE ATTRIBUTION OF SIMILARITY:

THE INFLUENCE OF SUCCESS AND FAILURE[1]

ANTHONY J. SMITH
University of Kansas

A new approach to the understanding of interpersonal relations has been provided by Heider (1958) in his conception of balanced states. Several recent investigations provide empirical support for Heider's ideas. They are consistent with the theory in their finding that in the triadic relationship of "p," "o," and "x," referring usually to the subject (p), another person (o), and an act or characteristic (x), respectively, subjects complete a partially defined triad in such a way as to produce cognitive equilibrium or balance (Smith, 1957). The magnitude of the force toward equilibrium has been shown to be a function of the valence for p of the two entities o and x (Smith, 1958). Furthermore, triads that are wholly defined in such a way as to provide balance are reported as being more pleasant than those which according to the theory are unbalanced (Jordan, 1953). In this instance, ratings of pleasantness are taken as crude measures of psychological equilibrium. Finally, group members in reporting their own sociometric choices and their predictions of others' choices yield results that conform to the theory of cognitive balance (Kogan & Tagiuri, 1958).

However, a question might be raised concerning the generality of the tendency toward balance. Are predictions based on the theory of balanced states to be supported in circumstances other than the essentially neutral contexts found in the experiments cited? Furthermore, what effect does an individual's experience of success or failure have on his tendency to attribute similarity to similar and dissimilar persons?

Schachter has reported a series of investigations concerning the need for affiliation (Schachter, 1959) based on earlier work by Festinger (1957). These studies in conjunction with the previously cited research provide a basis for speculating concerning the relationship between the two sets of variables. Schachter found that when persons were given the choice of being alone or being with others they preferred to be in the company of others. Furthermore, when given a choice of companions they tended to select those who were presumably most like themselves (with respect to participation in the experiment). Finally, under conditions of high anxiety induced by threat the need for affiliation with others was increased.

Schachter's first two findings can be related to the results of studies referred to earlier. When p's are asked to indicate their willingness to associate with various o's half of whom are similar, the remainder dissimilar, their average judgment corresponds to mild acceptance. Furthermore, the strength of acceptance is a function of the degree of similarity. Moderate acceptance is accorded similar o's whereas dissimilar o's are responded to with indifference (Smith, 1957).

Schachter's third finding makes it reasonable to assume that p's who experience failure should as a result of the threat involved also experience increased anxiety and have a greater need for affiliation. The increased need should in turn influence the valence of various o's. Festinger's observation that p's who experience moderate dissonance tend to avoid situations that would increase dissonance and seek situations that reduce it indicates that the initial valence of o is crucial. Similar o's who have been demonstrated to have positive valence for p would be expected to have higher valence following p's failure. The effect of failure on p's perception of dissimilar o's is equivocal. Both Heider's and Festinger's theories predict that dissimilar o's should have negative valence. Failure should in turn increase the negative valence. On the other hand, empirical evidence indicates that dissimilar o's are perceived to have neutral valence. In this

---

[1] This study was supported by a grant from the University of Kansas Research Fund. The author wishes to acknowledge the assistance of Ronald Sobol and John Park.

instance it is impossible to predict whether failure would have no effect on valence or would induce negative valence. Whether one follows the theory or the empirical evidence the differential in valence for similar and dissimilar o's should be greater for the subjects experiencing failure than for subjects experiencing success.

On the basis of these observations and the earlier finding that there is a positive relationship between valence and the attribution of similarity (Smith, 1958), the hypothesis is advanced that those p's who experience failure should show a greater differential in the attribution of similarity to similar and dissimilar o's than should p's who experience success.

It is also hypothesized that an earlier finding derived in a neutral setting to the effect that p should attribute greater similarity to similar o's than to dissimilar o's can be duplicated in a nonneutral setting.

Finally, though no specific hypothesis can be proposed, it is possible to determine whether success and failure have differential systematic effects on the tendency to attribute similarity to similar and dissimilar o's.

METHOD

Subjects

The subjects in this investigation, 128 undergraduate students enrolled in a course in general psychology, were drawn in equal proportion from each of two summer session classes and an intervening class in the spring semester. They did not volunteer but were required to participate in outside research as an integral part of their introduction to the field of psychology. The Allport-Vernon-Lindzey Study of Values was administered to all class members during the first meetings of the discussion sections of the course as part of their course content.

Procedure

Pairs of unacquainted like-sexed subjects were assembled 6 to 12 weeks after the administration of the test. They were informed that while psychologists had made significant progress in the study of the nonpersonal aspects of intelligence, e.g., numerical ability and spatial visualization, the relationship between such factors and the social expression of intelligence had been only partially explored. It was the purpose of this research to increase our understanding of this relationship. They were to be asked to work on four tasks, the first three of which were standard tests of intelligence taken from the Factored Aptitude Series (King, 1950). Their percentile scores would be revealed to them immediately following each test. It was suggested that their scores might be improved if they were to compete against one another.

Prior to the appearance of the subjects one was designated at random as the successful subject, the other as the failure. Following the administration of the first test (Judgment) the successful subject was told that he had received a percentile rating of 89, while the failure subject was told that his percentile rating was 28. On the second test (Dimension) they received percentile ratings of 85 and 30. The results reported to them for the third test (Memory) were percentile scores of 95 and 16, respectively. It was assumed that these patterns of performances would tend to induce in college students experiences of success and failure. The third task, Memory, was selected to provide a reasonable transition from the typical factored intelligence tests to a more socially oriented task.

Following this the subjects were given the final task. They were reminded that at best they always had limited information concerning other persons. On the basis of this information they were often required to infer other characteristics. The accuracy of such inferences was a measure of one form of social intellectual functioning.

Each subject was then given a partially completed Allport-Vernon-Lindzey test booklet which he was asked to study until he had developed a fairly clear impression of the person from whom the test information had been derived. Then he was requested to complete the test booklet as he believed the other person had. Each person was informed that this "other person" was actually his colleague.

In reality these experimental test booklets were derived from the subject's original test booklet. The partially completed test booklets were prepared so as to provide a systematic degree of similarity or dissimilarity. The method has been described elsewhere (Smith, 1957). Half of the subjects in the successful group were chosen at random to receive similar booklets. The remainder were provided with dissimilar test booklets. A similar procedure was used for the failure group.

In Part I of the test all information on any item involving the subject's highest value was excluded. In Part II for each such relevant item the rank assigned to this crucial value was excluded as was the rank two steps removed. By scoring this crucial scale on the experimental booklet it was possible to determine to what extent the subject attributed to the other person values that were like his own. The absolute difference between the subject's own score and that which he attributed to his colleague was taken as a measure of attributed similarity with smaller values representing greater attribution of similarity.

After these data had been collected the actual nature of the experiment was described to the subjects and each one was given the scores earned in the first three tasks.

RESULTS

Before examining the hypotheses, three preliminary analyses were undertaken.[2] It was neces-

---

[2] In each of the analyses to be reported it was neces-

TABLE 1

ANALYSIS OF VARIANCE OF ATTRIBUTION OF SIMILARITY SCORES

Source	df	MS	F
Success-Failure	1	342.89	5.94*
Similar-Dissimilar	1	1803.75	31.24**
Su-Fa × Si-Di	1	135.09	2.34
Within	124	57.73	
Total	127		

* $p < .02$.
** $p < .001$.

first determined that the original Allport-Vernon-Lindzey test scores on the crucial scales were homogeneous for the four experimental groups ($F = 1.96$; $p > .05$). The differences between experimenters and among replications were then tested in separate factorial analyses of variance. In each case main effects and interaction terms involving main effects were found to be insignificant.[3]

In the final analysis the data were arranged in a 2 × 2 factorial analysis of variance design with success-failure and similar-dissimilar as the main effects. The results of this analysis as presented in Table 1 reveal that success and failure have significant differential effects on the attribution of similarity ($p < .02$) with greater similarity being attributed by failure subjects. Moreover, the degree of similarity or dissimilarity of the partially completed test booklets is found to be related to a very significant degree to the amount of attributed similarity ($p < .001$) with greater additional similarity being attributed to the presumably similar colleagues. Finally, the interaction effect is found to be insignificant ($p > .15$).

These results are more clearly shown in Table 2 which presents the means for the various groupings. It can be seen that subjects in the failure group attributed greater similarity to their colleagues than did subjects in the success group for both the similar and dissimilar conditions. The differential between similar and dissimilar colleagues is 5.45 for the failure group as compared with 9.56 for the success group.

TABLE 2

MEANS OF DIFFERENCES BETWEEN OWN AND ATTRIBUTED SCORES FOR SIMILAR AND DISSIMILAR INDIVIDUALS UNDER CONDITIONS OF SUCCESS AND FAILURE

Conditions	Hypothetical Characteristics of Colleague		
	Similar	Dissimilar	Total Group
Success	8.38	17.94	12.84
Failure	7.17	12.61	9.94
Total Group	7.67	15.12	11.39
Previous Study	7.05	16.18	11.61

It is possible to get some crude idea concerning the locus of the effects of success and failure by comparing these results with those obtained earlier under neutral conditions (Smith, 1957). Exact comparisons are not possible because the earlier study used the same subjects for both similar and dissimilar conditions and did not involve the use of intelligence tests or success-failure. Under these neutral conditions (see Table 2) the means of the absolute difference between own and attributed scores for hypothetical similar individuals was found to be 7.05; for dissimilar individuals, 16.18, with a differential of 9.13.

A comparison of the three means obtained for the condition of similarity (8.38, 7.16, and 7.05) reveals substantial agreement. A comparison of the differentials shows very similar results for the success and neutral conditions (9.56 and 9.13) and an appreciably lower value for the failure condition (5.45). The effect of success and failure apparently occurs primarily in the failure condition as a result of the attribution of greater similarity to the dissimilar individuals (12.61) than occurs in the success condition (17.94). The difference between these two means is found to be significant ($t = 2.50$; $p < .02$). The success condition in turn is found to resemble the neutral condition (16.18).

DISCUSSION

In terms of Heider's theory, a subject who examines a similar test booklet can experience for each item the Condition pLx, oLx or $p \sim Lx$, $o \sim Lx$. In each instance there is a

---

sary to determine whether the assumption of homogeneity of variances could be met. This was tested by means of Hartley's procedure (Walker & Lev, 1953). In each instance the hypothesis of homogeneous variances was found to be acceptable.

[3] Tables summarizing these results are omitted in order to conserve space. The unreported results are essentially duplicated in the analysis of variance reported in Table 1.

tendency to attain balance in the pox triad. This is accomplished when pLo. Thus there is a cumulative effect in inducing a force corresponding to pLo. When a subject is required to predict his colleague's response, two elements are known: his own response to x and pLo. Balance is achieved when he attributes his own response to o (pLx, oLx or p ∼ Lx, o ∼ Lx). When, on the other hand, he examines a dissimilar test booklet, he observes for each item either pLx, o ∼ Lx or p ∼ Lx, oLx. Balance occurs when p ∼ Lo. When he estimates unknown characteristics of o, he can achieve balance by attributing to o characteristics other than his own (pLx, o ∼ Lx or p ∼ Lx, oLx). It should be noted, however, that if the lower rankings or weightings on the Allport-Vernon-Lindzey instrument represent indifference rather than dislike, it can only be predicted that the acceptance of dissimilar o's should be less than the acceptance of similar o's.

The first hypothesis to be examined in the present study states that the differential in attribution of similarity to similar and dissimilar individuals should be greater for subjects experiencing failure than for those experiencing success. An examination of the interaction term that reflects this relationship (Success-Failure × Similar-Dissimilar) gives no support to the hypothesis. In fact, the trend is in the opposite direction.

It should be recalled that this hypothesis was based on the assumption that the difference between the valences of similar and dissimilar persons would be increased following the experience of failure. The hypothesis further requires that the initial valence of the dissimilar person be neutral or negative.

The experiment was originally designed with the expectation that similar and dissimilar test booklets would completely define the degree of similarity. In retrospect, however, it can be seen that before any experimentally introduced evidence was available concerning similarity, each subject could easily observe that he and his colleague were either alike or indistinguishable with respect to many characteristics, e.g., sex, age, race, educational status. In each instance in which he perceived a shared characteristic (pUx, oUx), balance would be attained when the Condition pLo was established. Therefore following the introduction of dissimilar test booklets which ordinarily generate neutral valence, one would now expect a low degree of positive valence. If this is the case, failure should result in increased positive valence for the dissimilar subjects and no significant interaction between success-failure and the attribution of similarity would be expected.

From this line of reasoning one would also infer that the attribution of similarity should be greater for subjects experiencing failure than for those experiencing success. An examination of the significance of the main effect corresponding to success-failure indicates that the results are consistent with this inference. Moreover, the difference occurs for both similar and dissimilar individuals though only in the latter case is it significant.

Turning to the second hypothesis, the results indicate that in responding to hypothetical persons subjects attribute significantly greater added similarity to similar individuals than to dissimilar individuals. The t of 5.59 obtained in the present study compares with a similar finding (t = 6.12) reported in an earlier study (Smith, 1958). This finding provides empirical support for the hypothesis that predictions based on the theory of balanced states are confirmed not only in neutral contexts but in the nonneutral contexts corresponding to success and failure.

SUMMARY

Sixty-four pairs of unacquainted like-sexed subjects were given a series of three tests of intellectual ability. One subject in each pair was informed that he had earned high percentile scores on each task, while the other subject was told he had earned low scores. Each subject then received a partially completed Allport-Vernon-Lindzey test booklet designed either to be identical with or systematically dissimilar to the one completed by the subject 6 or more weeks earlier. *S* was asked to study this test booklet presumably derived from his colleague until a relatively clear conception of the person emerged. Then he was to complete the booklet as he believed his colleague had. Degree of similarity attributed was defined as the absolute difference between own and attributed score on a crucial scale.

The results revealed that failure subjects at-

tributed greater similarity to their colleagues than did successful subjects, particularly with respect to presumably dissimilar colleagues.

A previous finding in a neutral context involving the greater attribution of similarity to similar persons than to dissimilar persons was confirmed in the present nonneutral context.

The predictions that persons experiencing failure will show a greater differential in the attribution of similarity to similar and dissimilar individuals than will persons experiencing success was not supported. A possible explanation was offered.

REFERENCES

FESTINGER, L. *A theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

JORDAN, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6**, 273–287.

KING, J. E. *Factored Aptitude Series.* Chicago: Industrial Psychology, Inc., 1950.

KOGAN, N., & TAGIURI, R. Interpersonal preference and cognitive organization. *J. abnorm. soc. Psychol.*, 1958, **56**, 113–116.

SCHACHTER, S. *The psychology of affiliation.* Stanford: Stanford Univer. Press, 1959.

SMITH, A. J. Similarity of values and its relation to acceptance and the projection of similarity. *J. Psychol.*, 1957, **43**, 251–260.

SMITH, A. J. Perceived similarity and the projection of similarity: The influence of valence. *J. abnorm. soc. Psychol.*, 1958, **57**, 376–379.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

(Received September 8, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 424–431

THE INDIRECT REPRESENTATION OF SEXUAL STIMULI BY SCHIZOPHRENIC AND NORMAL SUBJECTS[1]

LEO B. AREY

Veterans Administration Hospital, Hines, Illinois

THIS study is concerned with the nature of the perceptual hypotheses advanced by schizophrenics and normals when confronted with sexual pictures presented at a speed just below the time threshold for conscious recognition and report. The method of tachistoscopic experimental attack was suggested, in part, by the work of Poetzl (1917), whose findings have been substantially confirmed by the recent studies of Shevrin and Luborsky (1958) and Fisher (1954, 1956). Poetzl's results indicated that stimuli apparently perceived outside of the focus of conscious awareness could later appear, symbolically transformed, in the manifest content of his subjects' dreams. One of the intriguing problems suggested by the Poetzl data was the temporal locus of these symbolic transformations. In the course of a preliminary experiment, directed toward the testing of a hypothesis of levels in symbolism, and involving the tachistoscopic showing of sexual pictures to schizophrenic and normal subjects, I observed an interesting phenomenon. In certain threatening areas the person responded to inadequately perceived sexual stimuli with perceptual distortions couched in images that were often remarkably similar to dream symbols described in the psychoanalytic literature. I observed this phenomenon many times before it finally occurred to me that I might be witnessing an immediate symbolic transformation of perception. Further, while these hypothetical symbolic responses of normal and schizophrenic subjects were often alike in quality, there were sometimes striking differences in the reported imagery.[2]

My observations led to the formulation of two general hypotheses that constitute the conceptual background of the current experiment:

1. When a person is confronted tachistoscopically with pictures representing various aspects of sexuality, the antecedents of objectively accurate conscious perception are often images that indirectly represent the stimulus object or activity.

2. These images reflect significant aspects of the subject's fantasies and affective attitudes concerning the stimulus object or activity, i.e., the kind of image that is used suggests part of the unconscious meaning the indirectly represented sexual object or activity has for the individual. If these images are considered as models of the way in which the subject unconsciously perceives sexual objects, it is then possible to judge whether his more active fantasies concerning these objects are such as to sustain and facilitate or to inhibit and obstruct his attempts to achieve satisfying interpersonal relationships and to maintain a relatively adequate integration of personality. An image can be rated as "sustaining" (healthy) if it reflects a fantasy of the sexual object or activity as a source of libidinal gratification and support; it can be rated as "nonsustaining" (pathological) if it reflects a fantasy of the sexual object or activity as destructive, threatening, or non-

---

[1] This paper is based on a doctoral dissertation submitted to the University of Chicago in 1952. The experimental work was carried out while the author was a psychology trainee at the Veterans Administration Hospital, Hines, Illinois. Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

Morris I. Stein, James G. Miller, Roy Brener, and David Shakow have all contributed invaluable advice, criticism, and moral support for which the author is deeply indebted. The writer is also grateful to Donald W. Fiske, who made many valuable suggestions concerning the statistical treatment of the results of this experiment.

Finally, this study owes much to the Psychiatry Service at the Veterans Administration Hospital, Hines, Illinois, for the cooperation of its staff in the selection of experimental subjects, and to the clinical psychology staff members and trainees at the Chicago Veterans Administration Mental Hygiene Clinic who served as judges in this experiment.

[2] This study is unpublished, but its design and results are described in Ch. I–II of my doctoral dissertation (Arey, 1952).

gratifying. The sustaining image will suggest that the subject's response to the object is essentially adient, with positively toned affective attitudes. The nonsustaining image will suggest that the subject's response to the object is predominantly abient, with fearful or negatively toned attitudes; or adient, with hostile or fearful attitudes. The characteristics that the subject attributes to the object may be projections of his own impulses toward the object, but the point is that these characteristics, whether they belong to the subject or the object, should somehow be expressed in the image used to represent the object.

The specific experimental hypothesis, in its most conservative form, reads: assuming that schizophrenics are more threatened by various aspects of sexuality than are normals, and assuming that the prerecognition distortions of tachistoscopically presented sexual stimuli are often indirect representations reflecting the subject's immediate affective response to the stimulus object, then a group of normal subjects can be differentiated from a group of schizophrenic subjects on the basis of the affective attitudes implied in the imagery of their respective prerecognition distortions.

The experiment was originally planned to deal only with the perceptual distortions that could be reasonably judged as "true symbols," but it became evident that it was going to be extremely difficult to assign limits to what constituted a "true symbol." In many instances it was impossible to differentiate between "true symbolism" and other forms of indirect representation, such as substitution of the part for the whole, allusion, plastic imagery, preconscious or conscious metaphor, etc.[3] Yet, excluding any of these distortions would have left too much room for the operation of unconscious biases in the selection of data. Consequently, both the experimental and the two general hypotheses, which were originally phrased in terms of "symbols" and "symbolism," are now stated in terms of "imagery" and "indirect representation." Although the judges were instructed to treat *all* the distortions as symbols, it is clear that they were dealing with the general class of indirect representations, and the interpretation of the experimental data must take this into account.

Another problem was posed by the distortions that might be mere misidentifications, with no symbolic or indirect representational reference. Some of these nonrepresentational responses to sexual stimuli were almost surely included in the data, but similar nonrepresentational responses to nonsexual stimuli provided a control for their effects.

METHOD

Design

The hypotheses were tested by using an experimental group of 25 schizophrenic male subjects and a control group of 25 normal male subjects, with the groups equated for age, intelligence, ethnic group, and education. Each subject was shown a series of 24 pictures representing various aspects of sexuality (sexual intercourse, birth, the genitalia, etc.) and he was asked to describe what he saw at .01-second presentations of the stimuli. As a control against the possibility that the responses given to these stimuli were not a specific function of the nature of the ideational areas represented, 9 blank cards and 18 nonsexual pictures were also presented.

Subjects

The experimental group consisted of 25 male schizophrenic subjects from the psychiatric wards at the Veterans Administration Hospital, Hines, Illinois. Criteria for inclusion were:

1. The psychiatrist in charge of the patient must express reasonable certainty that in his opinion the patient can be classified as schizophrenic, and another psychiatrist familiar with the case must confirm this diagnosis.
2. The clinical record must not reveal any suspicion of organic trauma or of mental defect per se.
3. The patient must be in sufficient contact to communicate with the experimenter and he must demonstrate during the presentation of the practice stimuli that he is cooperative enough to make an attempt to see the stimuli and to report on what he sees.

This last specification is quite important, for much of the psychological research on schizophrenia ends in ambiguity because the researcher often confuses a subject's refusal to participate in a task with an inability to perform it (cf. Bleuler, 1950, p. 58; Cameron, 1944, p. 903). Various internal controls indicate that a high level of cooperation was obtained from the subjects, who were mostly acute treatment cases rather than chronic, long hospitalized patients.[4] Further, no

---

[3] For an analysis of the criteria differentiating "true symbolism" from other forms of indirect representation, see Jones (1949).

[4] A check of the hospital records approximately 8 months after the collection of the experimental data revealed that seven of the group had been discharged with "maximum hospital benefits," seven were on extended trial visits, one had been discharged "against

TABLE 1

EQUATION OF EXPERIMENTAL AND CONTROL GROUPS

	Age in Years			IQ (as Measured by VIBS Scale)			Education (in Years of School Completed)		
	Range	Mean	σ	Range	Mean	σ	Range	Mean	σ
Normal (*N* = 25)	21–47	31.32	8.40	83–134	111.68	14.57	7–20	11.96	3.62
Schizophrenic (*N* = 25)	22–52	31.36	7.97	89–137	110.44	12.52	5–15	10.62	2.43
	$CR = .017$ $p = .98$			$CR = .319$ $p = .75$			$CR = 1.50$ $p = .13$		

patients were included who were currently receiving either electroshock or chemoshock therapy. It required an original starting group of 36 to provide the 25 subjects meeting the criteria for inclusion.

The control group consisted of 25 "normal" male subjects. None of these subjects was judged by the experimenter to be obviously severely neurotic or psychotic on the basis of a brief "neurotic inventory" (Clark revision of the Thurstone Personality Schedule) consisting of 25 items, supplemented by observations made while an abbreviated form of the Wechsler-Bellevue Intelligence Scale (VIBS) was being administered. However, there were probably several neurotic individuals and at least one latent schizophrenic included in the control group. These subjects were retained in the group because these opinions concerning them were formed on the basis of performance in the experiment itself, plus material confided to the experimenter after completion of the experimental task.

The groups were equated by first choosing the experimental group, and then selecting the control group so as to be equated with it in terms of age, intelligence, ethnic group, and education. Forty-one normal subjects were tested; five of them were excluded because the experimenter elicited enough evidence of neurotic traits to make the decision to exclude them before the experimental task was begun. In choosing the 25 control subjects from the remaining pool of 36 candidates, the experimenter tried to match the groups as closely as possible in terms of age and IQ (as measured by the VIBS scale). The equations for age and intelligence were quite satisfactory; that for education was less so, but the difference between the groups was not statistically significant. There were 20 white and 5 Negro subjects in each group.

Stimuli

First, there were 10 practice figures, presented for the purpose of acquainting the subject with the use of the tachistoscope and assuring the experimenter that the subject was likely to try to cooperate. These figures were simple designs, e.g., plus sign, circle, triangle, square, arrow.

The test and control stimuli, described in order of presentation to each subject, were:

1. A woman's lips, smiling, with teeth showing[5]
2. Blank card
3. Anatomic drawing, showing a fetus in the mother's womb
4. Blank card
5. A young woman's breasts
6. Blank card
7. The painting by Barend van Orley, "Neptune and the Nymph," an unmistakable representation of two nude figures in heterosexual intercourse
8. Blank card
9. A man's lips, open slightly
10. A cedar chest
11. Rodin's "The Thinker"
12. A little boy, standing, dressed in sleeping suit
13. A doctor, holding a newborn baby upside down
14. A modern house, with wide balcony and straight, overhanging eaves
15. A wood carving, showing a young man masturbating. The figure is shown from the chest to about the middle of the thighs.
16. A formal garden, with paths leading to a central pool
17. A view of Rodin's "The Kiss," chosen to give an impression of heterosexual intercourse
18. A woman smiling, playing with and partly holding a happy looking baby, who is sitting on a bed
19. The male genitalia. The penis is erect.
20. A dining room table. The table is bare except for a bowl of flowers in the center.
21. A woman lying on her back, with legs outspread and with genitalia clearly shown

---

medical advice," three had been discharged as "absent without leave," three had been transferred to other hospitals, and four of the group were still in the same hospital. Such statistics are difficult to interpret, but they do tell us that 18 of the 25 experimental subjects could be temporarily absorbed at some level by the society outside the hospital within a relatively short time.

[5] Three pictures of an "open mouth" (No. 1, 9, 45) were included in the series for exploratory purposes only, and responses to these stimuli were excluded from the experimental data. No. 45 was a duplicate photograph of the No. 1 stimulus. With this one exception, all the pictures were different despite the identity of some descriptions. Unless otherwise specified, e.g., as a drawing or a painting, the figures were photographs of the subject matter described.

22. The heads of two dogs, one above the other
23. A drawing of the male genitalia. The penis is erect.
24. A chair next to a small table on which there is a desk lamp
25. A young woman's breasts
26. The face of a young, smiling child
27. The nude torso of a young woman, lying down
28. Blank card
29. A painting of Hercules, showing a very muscular man with arms drawn back ready to strike at something or someone
30. A young colt
31. A woman's body, from the chest down. Her legs are spread open, with the genitalia clearly shown.
32. Drawing of three interlocking circles, with the interiors of the circles shaded by pencil markings
33. A view of Rodin's "The Kiss." This can be seen as either a kiss between two nude people or it can be seen as heterosexual intercourse.
34. A small white country church with a black roof, topped by a black steeple supporting a white cross
35. Drawing of the male genitalia. The penis is flaccid.
36. Interior corner of a living room, showing couch, chair, lamp, etc.
37. A young woman's breasts
38. Approximately a dozen books, lined up side by side in a vertical position
39. A statue of a nude youth holding a small, elongated jar just above and in front of the genital area. He is pouring water into his cupped left hand.
40. The face of a beautiful young woman with flowing hair
41. Nude torso of a young woman, with quite prominent breasts
42. A young man sitting on a couch; he has turned to face a young woman standing behind the couch. The woman is offering him a dish of fruit.
43. A Pompeiian bronze; the statue is called "Die Potenz." The head is bent, slightly forward, as the man gazes at his erect penis.
44. A panoramic view of rolling slopes and softly contoured hills
45. A woman's lips, smiling, with teeth showing
46. Photograph of a nest of suitcases. A large suitcase is opened; in it is a smaller suitcase, opened; in it, in turn, another even smaller suitcase is opened. Altogether, there are six of the suitcases.
47. A nude woman, from her crotch to her feet. She is sitting on the edge of a bed, with her legs spread wide.
48. Blank card
49. Drawing of a nude male figure from chest to middle of thighs. His hand grasps his erect penis.
50. Blank card
51. Close-up of the female genitalia. A small doll's head and body protrude from the vaginal opening. At fast tachistoscopic speeds, this picture provides an excellent representation of birth.
52. Blank card
53. Photograph of a nude female torso, with quite prominent breasts
54. Blank card

The ordinal number of presentation was even for all of the control stimuli and odd for all of the test stimuli. These pictures were photographed so that one dimension was always approximately 7 inches; the other dimension, which could be either height or width, was allowed to vary from 3.75 to 7 inches. They were mounted on 10.5 × 8.5 inch white carboard and could be inserted directly into the tachistoscopic viewing field. The nine blank controls were simply white cards with no pictures mounted on them.

Strictly speaking, all of the 18 nonsexual pictures have a potential symbolic sexual reference. But that sexual reference is certainly not the direct and straightforward one of the 24 frankly sexual stimuli.

Apparatus and Procedures

A Harvard (modified Dodge) Tachistoscope was used in the experiment. The exposure field was approximately 23 inches from the subject's eyes. The pre-exposure field and the exposure field were each illuminated by two white, 4-watt fluorescent bulbs, giving a brightness of 2.8 apparent foot-candles, as measured at the eyepiece of the machine. This illumination was then decreased by four sheets of 20-pound white bond paper placed over each of the four milk screens.[6]

Each subject was tested individually in a small room, with only the experimenter and the subject present. Each subject was instructed:

> Tell me exactly what you see when the picture flashes on in the machine. Even if you are not sure what the picture actually shows, I want you to tell me what your first impression of it is. It is very important that you tell me your impression every time a picture flashes in the machine.

The exposure time for every tachistoscopic presentation was always .01 second. Some of the stimuli were presented several times to subjects in both groups when the experimenter felt that especially significant material could be elicited. Also, some of the pictures correctly recognized by the subject on the first presentation were occasionally re-presented several times so that the subject would not believe that repeated presentation meant that his first response was wrong. All responses were recorded as nearly verbatim as possible. This was fairly easy to do, as most responses were rather brief. Only the responses to the first presentation for each stimulus were considered in the statistical treatment of the data. After the completion of the presentations for each of the first 15 cards, it was shown at unlimited time and the subject was asked for a description of it. The purpose of this procedure was to convince the subject that he need have no hesitation in reporting his perception of a sexually toned stimulus picture when he consciously recognized it as such.

At the conclusion of the experimental task, each subject was given the opportunity to see all the pictures in the series at unlimited time. Every subject in both groups took advantage of this opportunity. Each subject was then given the chance to talk about the pictures and to verbalize some of the feelings that the experimental task had generated.

---

[6] Recognition of tachistoscopically presented stimuli depends on a complex interaction of speed of presentation, brightness of the stimulus field, and the structural qualities of the stimulus. My experience indicates that the phenomenon of indirect representation is best elicited when about 10–20% of the subjects achieved accurate recognition of the stimulus, thus assuring that partial cues concerning the nature of the stimulus are likely to be accessible to the subjects who do not consciously recognize the picture.

Treatment of Data

Six independent male judges (consisting of four members of the clinical psychology staff and two advanced VA clinical psychology trainees, all at the Chicago VA Mental Hygiene Clinic) rated the responses given to several of the sexual stimuli. Four judges rated the responses to three of the pictures; one judge rated the responses to four of the sexual pictures; the sixth judge rated the responses to five of the sexual pictures. The experimenter's own ratings of the responses as to imagery level clearly indicated that it was safe to assume that if the judges could differentiate the two groups on the basis of the responses to these few sexual pictures, then they should be able to make differentiations at as high, or higher, levels of confidence by using the responses to the full set of 24 sexual stimuli.

These judges had only the actual responses, with no way of determining whether a response was given by a schizophrenic or a normal except by a judgment of the response itself. The general instructions to the judges were as follows:

> These are the responses given by 25 normals and 25 schizophrenics to this picture when it was presented to them tachistoscopically at .01 second. [The specific picture was then placed before the judge.] There are some recognitions, and there are some gross perceptual distortions and misidentifications of the stimulus picture. The recognitions are to be grouped in one category. Even if you can tell from language cues, or what have you, that the response was given by a schizophrenic, if the response is essential recognition of the central concept of the picture, then it goes into the recognition category. Assume that the perceptual distortions are symbols representing the way the subject unconsciously perceives the object, or the way in which he unconsciously tends to respond affectively to the object. Assume that they are the models that he unconsciously lives by. Assuming this, try to decide whether the subject's impulses toward the object are libidinally or destructively cathected. If the symbol is a libidinal one, indicating that the object is perceived as a source of gratification, then rate the response as "sustaining." By this we mean that this kind of model would tend to sustain the person's attempts at satisfactory interpersonal relationships, as well as tend to sustain a relatively satisfactory level of personality integration. If the symbol is a destructive-fearful one, indicating that the object is feared, hated, or perceived as essentially nongratifying, then rate it as "nonsustaining." By this we mean that this sort of model would tend to interfere with or obstruct attempts at satisfactory interpersonal relationships and the maintenance of a relatively satisfactory level of personality integration.

Only these very general instructions were given to judges for rating Pictures 7, 17, 33, 47, and 51. For Pictures 15 and 43, more specific instructions were given.

For 15 the additional instructions were as follows:

> To be rated as sustaining, the symbol must have some formal similarity to the idea or object symbolized. It must suggest an intact penis. To be rated as nonsustaining, the symbol may express either the idea of castration (as a consequence of masturbation), or it may express an idea of violence beyond the control of the subject, i.e., of forces beyond the subject's control. Also include here all feminine identifications, or symbols suggesting feminine identification of the stimulus (assuming that feminine identification implies castration).

For Picture 43, "Die Potenz," the additional instructions were as follows:

> Assume that the symbol reflects the subject's unconscious phantasy concerning his potency, his unconscious acceptance or nonacceptance of potent masculinity. Will such a phantasy sustain or will it obstruct his attempts at satisfactory interpersonal relationships? Rate the perceptual distortions accordingly.

After all the ratings had been completed and recorded, the two respective imagery scores accruing to each subject on the basis of the judge's ratings were tabulated. Since the hypothesis demands that the two groups of subjects be differentiated by the respective kinds of indirect representation they use for sexual objects and activities, the responses showing adequate recognition and the responses of stimulus rejection (e.g., "Don't know what it is," or "My eyes blinked then," etc.) obviously could not be included in the tabulation of imagery scores. The resulting quantitative difference in the scorable distortions produced by the two groups introduced a source of error in interpreting the judgments. For instance, suppose that in rating responses to five pictures, a judge found that 110 responses from the schizophrenic group were scorable either as sustaining or as nonsustaining images, but that only 100 responses from the normal group were scorable images, with the remainder being rated as recognition or stimulus rejection. Chance factors alone would then favor the schizophrenics' obtaining higher scores than the normal subjects in both categories of imagery. To take care of this difficulty, a specific "correction constant" was derived for each judge, thus: Let x equal the sum of the mean of the sustaining imagery scores and the mean of the nonsustaining imagery scores assigned by the judge to the schizophrenic group on the basis of the responses to the pictures rated by him. Let y equal the sum of these two means for the normal group. Then, the constant, c, is found by solving the equation: $cy = x$. The means of the normal group's respective scores for sustaining imagery and for nonsustaining imagery were then multiplied by the obtained constant. It was assumed that the standard deviations of the two distributions were not changed. This device permitted a direct comparison between the means of the schizophrenic group and the normal group for each of the two kinds of imagery scores.

TABLE 2

DIFFERENTIATIONS BETWEEN 25 SCHIZOPHRENICS AND 25 NORMALS MADE BY SIX JUDGES IN ASSIGNING NONSUSTAINING IMAGERY SCORES

Judge	Code numbers of respective pictures for which each judge rated the responses	Means of nonsustaining imagery scores assigned by each judge to the two groups		t
		Normals	Schizophrenics	
F	15, 33, 43	1.01 (0.82)[a]	2.08 (0.74)	4.75***
D	7, 15, 33, 43	1.76 (0.99)	2.84 (1.08)	3.62***
E	7, 15, 43	.66 (0.63)	1.40 (0.80)	3.54***
A	7, 15, 43	1.45 (0.78)	1.92 (0.93)	1.88
B	15, 33, 43, 47, 51	1.80 (1.02)	2.24 (1.14)	1.42
C	7, 17, 33	1.77 (0.75)	1.84 (0.88)	0.30

Note.—Judges F, E, A, C rated responses to three pictures. Thus in the rating of these judges, the highest possible score an individual can accrue is 3. In Judge D's ratings of four pictures, 4 is the ceiling for an individual's score; in Judge B's ratings, 5 is the ceiling.

Mean score of normal group is corrected by a specific constant for each judge in order to rule out the effect of the schizophrenics' production of more scorable distortions as a factor in the judges' discriminations between the two groups (see text). The uncorrected normal means: Judge F, .96; Judge D, 1.76; Judge E, .64; Judge A, 1.28; Judge B, 1.64; Judge C, 1.40.

[a] Standard deviations in parentheses.

*** $p < .001$, two-tailed test.

TABLE 3

DIFFERENTIATIONS BETWEEN 25 SCHIZOPHRENICS AND 25 NORMALS MADE BY SIX JUDGES IN ASSIGNING SUSTAINING IMAGERY SCORES

Judge	Code numbers of respective pictures for which each judge rated the responses	Means of sustaining imagery scores assigned by each judge to the two groups		t
		Normals	Schizophrenics	
F	15, 33, 43	1.55 (0.89)[a]	.48 (0.64)	4.78***
D	7, 15, 43, 33	1.76 (0.95)	.68 (0.84)	4.18***
E	7, 15, 43	2.10 (0.87)	1.36 (0.79)	3.07**
A	7, 15, 43	1.27 (0.71)	.80 (0.80)	2.15*
B	15, 33, 43, 47, 51	2.60 (0.98)	2.16 (1.22)	1.37
C	7, 17, 33	.55 (0.70)	.48 (0.49)	0.41

Note.—Same as note to Table 2, except that the uncorrected normal means here are: Judge F, 1.48; Judge D, 1.76; Judge E, 2.04; Judge A, 1.12; Judge B, 2.36; Judge C, .44.

[a] Standard deviations in parentheses.

* $p < .05$, two-tailed test.

** $p < .01$, two-tailed test.

*** $p < .001$, two-tailed test.

RESULTS AND DISCUSSION

Schizophrenic subjects were expected to give more nonsustaining images for sexual objects and activities than would normal subjects, and, conversely, normal subjects should use more sustaining images than should schizophrenics. The results shown in Table 2 and Table 3 confirm the prediction. All six judges assigned more of the schizophrenics' than of the normals' distortions to the nonsustaining category; conversely, all six judges assigned more of the normals' than of the schizophrenics' distortions to the sustaining category.[7] The differentiation between groups was in the predicted direction on both measures for all six judges, and the differentiation between groups was significant beyond the .01 level on both measures for three of the judges.

Two of the judges who differentiated between schizophrenics and normals at high levels of significance in rating the perceptual distortions to the sexual stimuli were asked to rate the responses to a large sample (11) of the nonsexual stimuli. They were unable to differentiate between the two groups at statistically significant levels on the basis of these ratings. For one judge, $.10 < p < .20$; for the second judge, $.30 < p < .40$.

---

[7] For Judge C, this is true for sustaining imagery scores only when the Normal mean is corrected to rule out the effect of the schizophrenics' production of more scorable distortions than are given by the normals.

As a background against which to evaluate the ratings of the six independent judges, my ratings of distortions given to the full set of 24 sexual stimuli yielded these differentiations in the predicted directions between normals and schizophrenics: on sustaining imagery scores, $t = 5.78$, $p < .001$; on nonsustaining imagery scores, $t = 5.26$, $p < .001$.

In rating distortions to Pictures 15, 33, and 43 (i.e., those rated by Judge F), my differentiations between groups were: on sustaining imagery scores, $= 5.86$, $p < .001$; on nonsustaining imagery scores, $= 3.90$, $p < .001$.

In rating distortions to Pictures 7, 17, and 33 (i.e., those rated by Judge C), my differentiations between groups were: on sustaining imagery scores, $t = 2.66$, $p < .02$; on nonsustaining imagery scores $t = 1.92$, $p < .1$. Judge C was given his particular group of pictures (all representations of heterosexual intercourse) to see if he could pick up any differences I might have missed in my ratings.

Thus Judge C was given a more difficult task than Judge F, for he was asked to make ratings on the responses to a group of pictures that in my own ratings yielded questionable, borderline differentiation. Judge F, on the other hand, rated responses to pictures that in my ratings yielded differentiations roughly equivalent to those obtained in rating the whole series of 24 sexual stimuli.

Judges D, E, A, and B also were given groups of pictures to rate for which my respective differentiations were roughly equivalent to those obtained in rating the whole series.

There were no differences between the two groups in response to the blank cards. No subject in either group reported seeing anything except a "blank," or a "flash," or a "light" when these cards were presented. One schizophrenic subject did respond "just a flash" to one card and then go on to tell a long story about Flash Gordon, but this narrative was clearly not intended as the description of a perception.

One of the assumptions of the experimental hypothesis is that most people in this culture are threatened by some aspects of sexuality, but that schizophrenics are more threatened by various aspects of sexuality than are normal persons. This assumption was supported by the recognition scores made by the two groups of subjects for the two classes of pictures.

The sexual and the nonsexual stimuli were rated for structural difficulty by two husband-and-wife teams, and it was possible to use their ratings to pick out 13 of the sexual pictures whose structural difficulty was considered as roughly equivalent to that of 13 of the nonsexual stimuli. All responses to both sexual and nonsexual stimuli were rated for recognition versus nonrecognition by the experimenter on the basis of criteria established before any of the data were collected.[8] On these two groups of equated pictures, the normal group recognized significantly more of the sexual pictures than did the schizophrenic group (t, 3.63; df, 48; $p < .001$). Normal subjects recognized slightly more of the nonsexual stimuli than did the schizophrenics, but the difference was not significant (t, 1.33; df, 48; $p < .20$). Further, both normal and schizophrenic subjects recognized significantly more of the nonsexual stimuli than of the sexual stimuli ($p < .001$). Finally, the differences in recognition scores for the two classes of pictures were significantly greater ($p < .02$) for the schizophrenic group (mean difference, 4.56) than for the normal group (mean difference, 2.96).

The differences in stimulus rejections by the two groups for the two classes of pictures parallel the differences in recognition. Schizophrenics exceeded normals in stimulus rejections of the sexual pictures ($p < .01$); in fact, none of the normals gave more than one stimulus rejection, while 14 of the schizophrenics gave two or more. Further, schizophrenics gave significantly more stimulus rejections in response to sexual stimuli than to the nonsexual stimuli ($p < .001$).

The differences in recognition scores of the two groups for the two classes of pictures support the assumption that male schizophrenics are more threatened by various aspects of sexuality than are normal young men.[9] While it cannot be experimentally demonstrated that reported recognition accurately corresponds to conscious recognition, it nevertheless seems reasonable to assume that differentials with respect to *report* of recognition are valid measures of differences in the experience and handling of sexual threat.

SUMMARY AND CONCLUSIONS

On the assumption that male schizophrenics are more threatened by various aspects of sexuality than are normal males, and that the prerecognition distortions of tachistoscopically presented sexual stimuli are often indirect representations reflecting the subject's immediate affective response to the stimulus object, it was hypothesized that a group of normal subjects could be differentiated from a group of schizophrenic subjects on the basis of the *affective attitudes* implied in the imagery of their respective prerecognition distortions.

This hypothesis was tested by using an experimental group of 25 schizophrenic male subjects and a control group of 25 normal male subjects, with the groups equated for age, intelligence, ethnic group, and education.

---

[8] Using these criteria, the experimenter's wife made independent judgments of recognition versus nonrecognition for the total of 2,100 responses from the two groups of subjects. There was less than 3% disagreement with my scoring, with the disagreements showing no consistent directional bias in favor of either the schizophrenics or the normals.

[9] Since only males were used as subjects in this experiment, the conclusions based on these data cannot be generalized to women. In this connection, it is of interest to note that Rosenstock (1951) has presented some experimental evidence indicating that women tend to repress aggressive material, while men tend to repress sexual material.

Each subject was shown a series of 24 pictures representing various aspects of sexuality (sexual intercourse, birth, the genitalia, etc.) and he was asked to describe what he saw at .01-second presentations of the stimuli. As a control against the possibility that the responses given to these stimuli were not a specific function of the nature of the ideational areas represented, 9 blank cards and 18 nonsexual pictures were also presented. The prerecognition distortions given to samples of both sexual and nonsexual stimuli were rated by independent judges as either "sustaining" or "nonsustaining" images, with the judges blind as to whether a response was given by a schizophrenic or a normal individual.

The quantitative experimental results support the following conclusions:

1. While both normal males and schizophrenic males in our culture are threatened by some aspects of sexuality, schizophrenics are threatened by more aspects of sexuality than are normals, and this differential in threat is not simply a function of the schizophrenics' indiscriminate fear of all kinds of objects and activities.

2. The indirect representation of sexual stimuli often occurs when these stimuli are tachistoscopically presented just below the time threshold for conscious recognition and report.

3. There are discriminable differences between normal males and schizophrenic males in the indirect representation of sexual stimuli under the conditions of this experiment. The schizophrenic group significantly exceeded the normal group in the production of nonsustaining imagery (reflecting negative affective attitudes) in response to sexual stimuli, while normal subjects significantly exceeded schizophrenics in the production of sustaining responses (reflecting positive affective attitudes) to these stimuli.

REFERENCES

AREY, L. B. A study in perception and symbolism. Unpublished doctoral dissertation, University of Chicago, 1952.

BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: International Univer. Press, 1950.

CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 2. New York: Ronald, 1944. Pp. 861–921.

FISHER, C. Dreams and perception. *J. Amer. Psychoanal. Ass.*, 1954, **2**, 389–445.

FISHER, C. Dreams, images and perception: A study of unconscious-preconscious relationships. *J. Amer. Psychoanal. Ass.*, 1956, **4**, 5–47.

JONES, E. The theory of symbolism. In E. Jones, *Papers on psychoanalysis.* (5th ed.) Baltimore: Williams & Wilkins, 1949. Pp. 87–144.

POETZL, O. Experimentelle erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Zbl. ges. Neurol. Psychiat.*, 1917, **37**, 278–349.

ROSENSTOCK, I. M. Perceptual aspects of repression. *J. abnorm. soc. Psychol.*, 1951, **46**, 304–315.

SHEVRIN, H., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.

(Received September 14, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 432–439

THE UNCONSCIOUS PERCEPTION OF THE MEANING OF VERBAL STIMULI[1]

MARCUS J. FUHRER AND CHARLES W. ERIKSEN
University of Illinois

WITHIN the last 10 years, a number of experiments (Blum, 1955; Dixon, 1956, 1958; Lazarus & McCleary, 1951) have attempted to demonstrate the operation of a supersensitive perceptual system functioning below the individual's awareness threshold. At the speculative level, Osgood (1957) has suggested that "certain dominant components of the mediating reaction to a sign can be elicited at stimulus levels below that at which other components occur" (p. 396). "Dominant evaluative (emotional) components" are included by Osgood in that class of mediating responses which may occur temporally prior to accurate recognition of the stimulus (p. 397). Lazarus (1956) has indicated that the dimension of threat versus nonthreat may be one of these subliminally discriminated, evaluative components.

Eriksen (1956a, 1956b, 1958) has formulated an alternative theoretical model which assumes neither prior nor supersensitive discrimination, but which accounts for correct discriminative responses in the absence of verbalized discrimination. The theory conceptualizes perception as a mediating process, occurring between stimulus input and response output. The perceptual process is assumed to give rise to a number of different responses, of which verbal identification is only one. As applied to subception-like effects, the model presupposes concurrent responses to a stimulus, e.g., a verbal response and an affective response, both of which are less than perfectly correlated with the stimulus and which possess independent sources of error. In response to impoverished stimulation, cases thus may arise in which the affective response is correct when the identification is inaccurate and vice versa. In contradistinction to the theories of Osgood and Lazarus, this formulation neither assumes that the affective responses occur temporally prior to verbal identification nor that the threshold for affective discrimination is lower than that for verbal identification. The only necessary conditions for obtainment of a subception-like effect is that both the verbal identification and the affective responses are responding to the mediating perceptual process and have at least partially independent sources of error.

The concurrent response model and the theorizing of Osgood and Lazarus have recently been put to a differential test by Eriksen, Azuma, and Hicks (1959). In this experiment, two sets of words, previously judged to be pleasant or unpleasant, were presented tachistoscopically to a group of subjects (*S*s). Following exposure of each word, the *S*s judged whether the word was pleasant or unpleasant and also guessed what the word was. The judgments of affective classification that occurred in the absence of correct identification of the word then were examined for above chance accuracy. The authors report that in the absence of correct identification of the stimuli, the *S*s' affective judgments were correct significantly more often than could be attributed to chance. However, as opposed to the theories of Osgood and Lazarus, no evidence was found that the affective judgments had a lower discrimination threshold than the identification responses.

An experiment by Dixon (1958) is of particular interest because it generated evidence of unconscious discrimination which appears relatively immune from alternative interpretation by the concurrent response model. Dixon hypothesized that subliminal stimuli would activate autistic verbal behavior which could be related meaningfully to the subliminal stimuli. Dixon presented a series of predominantly verbal stimuli at an illumination level so low that the *S*s reportedly were unable to discriminate even the presence of the stimuli. The *S*s' task was to guess the identity of the stimulus items. That the guesses were actually associations to the stimuli was shown

---

[1] This research was supported by Research Grant M-1206 from the National Institute of Mental Health, United States Public Health Service.

by the *S*s' better than chance ability subsequently to indicate the stimulus item to which each guess had been given. Because of these surprising results, and because Dixon explicitly labels his experiment a "pilot study," the present investigation was initiated to explore the possibilities inherent in Dixon's interesting method and possibly to extend his findings.

METHOD

Design

Four treatment groups were employed in the present investigation. In all groups, *S*s were run individually and participated in two separate sessions. The first session was divided into two stages: threshold determination and subsequent presentation of the stimuli. In the second session, held 3 days later, each *S* was presented with his responses made during the first session and asked to match these against the original stimulus items. Procedural variations among the four groups differed primarily in terms of how the thresholds were obtained, the level of illumination at which the stimulus items were presented, and the number of items to which the *S* matched his original responses.

The procedure used in Group I was directly comparable to that used by Dixon, except that GSRs were not recorded. The instructions also were changed slightly to increase the probability that the *S*s would not hesitate to report if anything was perceived during the presentation of the stimuli. While Dixon employed only 7 *S*s, 16 *S*s were used in our Group I.

To obtain threshold determinations for *S*s in Group II, a forced-choice technique was used. This technique has been shown by Blackwell (1953) to yield more reliable threshold values than techniques of the kind used by Dixon. Whereas the stimuli for Group I were presented well below the *S*'s previously determined threshold, the same stimuli were presented for Group II slightly above the obtained thresholds. It was anticipated that comparison of the results from Groups I and II might reveal a range of illumination values at which autistic responses to stimuli could be obtained.

Group III was employed to determine how much the *S*s' responses were influenced by structural characteristics of the stimuli, e.g., the height of the letters and the length of the words, as opposed to the meaning of the stimuli. For *S*s in Group III, a threshold was obtained such that the *S* could accurately identify approximately 15% of a series of test words. To determine the relative influence of structural cues, the *S*s matched their guesses to a list containing the actually presented stimulus items, as well as 10 other words of similar structural composition, but of reversed emotional connotation. For example, if the actual stimulus was "male sex organ," the *S*'s list included this phrase, as well as the affectively more neutral phrase "made six orders." Emotionally neutral stimuli also had structurally similar, but emotionally connoted counterparts.

A fourth treatment group was employed primarily to determine how response biases affect the *S*s' matchings under the conditions used in Group III. The stimuli were presented upside down and backwards, thus ruling out the possibility of an *S* perceiving the meaning of the verbal stimuli. By this method of presentation, structural cues were confounded, with the exception of word or phrase length.

Subjects

The *S*s were undergraduate students from the introductory psychology course at the University of Illinois. They were required to serve a minimum number of hours as experimental *S*s as part of their course requirements. Each of the 64 *S*s was randomly assigned to one of four treatment groups, consisting of 8 males and 8 females.

Apparatus

To minimize uncontrolled cues from the experimenter (*E*), the *S* was seated in a wooden booth 6 ft. by 6 ft. by 4 ft. The only openings were a door and an aperture 11 in. by 7 in. in the far wall 38 in. from the floor. The aperture, through which the stimuli were projected, was covered with a ground glass screen to give a homogeneous field illuminated with a brightness of .3 log ft-L. A chair was provided, as well as an adjustable chin-rest 38 in. from the floor and positioned 44 in. from and directly in front of the screen.

The stimulus items were photographed on 35-mm. film and placed on individual slides, masked except for the exposed portion of the film. An automatic projector, supplied by a constant voltage transformer through a variac, was used to project the stimuli. Image brightness was controlled by a specially calibrated projector diaphram, making increments of .05 log ft-L possible throughout the range of illumination used in the experiment. A shutter, manually operated by *E*, was used to expose the stimuli. Opening the shutter automatically activated a chronotron; closing the shutter automatically stopped the chronotron.

A standard set of stimuli were used in the four treatment groups. There were four "emotionally neutral" words and six words of "primary emotional interest." The items were presented in the following order: this, penis, barn, father, vagina, seven, male sex organ, line, mother, female sex organ. For Group I, this standard set of items was increased to include a horizontal and a vertical line, positioned second and fifth, respectively, in the list. For Group II, the stimulus item house was used to make threshold determinations. For Groups III and IV, threshold determinations were made using the items: country, time, other, almost, present, because.

Procedure

The procedure described for Group I was the same in each of the other three groups, with the exception of specific variations described below in detail. All *S*s were run individually in two separate sessions. The first session was divided into two stages, threshold determination and subsequent presentation of the stimuli.

Group I. The *S* was seated in the darkened booth and instructed to relax during an 8-min. period of dark adaptation. A clearly visible horizontal line then was projected onto the center of the screen, so as to subtend a visual angle of 2° 48′. The *S* was asked to position his chin on a chin-rest, and these instructions were read:

> You see before you, near the center of the screen, a horizontal line of light. I am going to make the line gradually dimmer. All you have to do is to watch it carefully and say "Stop" the moment it has completely disappeared. Don't say "Stop" until you are quite certain you cannot see anything at all upon the screen.

The brightness of the line was diminished at the rate of .1 log ft-L/sec., until the *S* said "Stop." A series of judgments was obtained also using a vertical line. The *S*'s threshold was defined as the lowest illumination level which was not equaled or bettered in three subsequent trials, considering both series of judgments.

Introductorily to the presentation of the stimuli, the *S* was instructed: I am going to project words on the screen but often they may be too dim for you to see. With each projection, however, I want you to make a guess regarding identity of the word shown. Your signal that a word is there and that you should respond will be my saying "Now." During the experiment, keep watching the screen even though you may see little upon it. Try and maintain a relaxed, passive attitude, just allowing each guess to be the first word that comes to mind after I say "Now." I will give you no indication of whether you were right or wrong.

Ten practice responses were obtained, without a stimulus word being presented. The projector diaphram then was adjusted to give a level of illumination .3 log ft-L below the *S*'s previously determined threshold. Simultaneously with the opening of the shutter, *E* signaled "Now," the timer was automatically started, and the first stimulus word was presented. Immediately following *S*'s response, the shutter was closed, and the timer simultaneously stopped. The *S*'s response latency and verbal response were recorded by *E*. The *S* was asked if he had seen anything on the screen. This procedure was followed for the presentation of each of the 12 stimuli. The stimuli were presented in the same order three more times, without indication that the series was being repeated, until four responses were obtained to each of the 12 stimuli. The *S* then was removed from the booth and was asked to associate to each of the responses he had given.

A second session was held 3 days later. A card was handed the *S* upon which were typed in random order the stimuli used in the first session. The *S* was asked if the list seemed familiar in any way. The following instructions then were read:

> I am going to read to you a list of words. After each I want you to tell me the number of that item on your list which in your opinion is most appropriate to the word I have given. I shall not ask you for the reasons for your choice which should be entirely personal and not bound by logical or common-sense considerations; just choose the item which you feel, for any reason whatever, is most suited to the word I have given. Take as much time as you like, then give me the number of your choice.

The *S*'s original responses were read in random order, and the *S*'s matches were recorded.

Group II. The *S*'s threshold was determined by a forced-choice technique. The following instructions were given:

> I am going to present a series of two presentations on the screen. In one of the two presentations, I will present nothing, while in the other, I will present a dimly illuminated word. Your task is to guess whether the word appeared in the first or the second presentation. I will say "the first" to signal the first presentation, pause 4 sec., and then say "the second" to signal the second presentation. I then will ask you for your guess. I will give you no indication of whether you were right or wrong. We will proceed then to the next set of presentations in the series.

A series of presentations was given; the position of the word *house* was randomly ordered within each set of presentations. The level of illumination was increased at increments of .2 log ft-L until *S* made seven out of eight correct judgments. This illumination level then was used to present 10 of the 12 stimuli used in Group I, the vertical and horizontal lines being excluded. In all other respects, the procedure did not vary from that used in Group I.

Group III. The *S*s' thresholds were obtained using the following instructions:

> I am going to present a series of different words on the screen. Your task is to guess the identity of the word shown. When I give the signal "Now," a word will be presented, and you are to announce your guess aloud. In some cases, the word may appear so dimly that you feel your guess is rather arbitrary. Don't let that disturb you, for you may be guessing more accurately than you think. I will give you no indication whether your guess was right or wrong.

The series of test words was repeated in random order at increments of .2 log ft-L until the *S* could accurately identify one of the seven words. At this illumination level, the standard set of stimuli were presented, with the prior instructions given:

> Your task remains the same, i.e., to guess the identity of the words presented on the screen. Don't be disturbed if you feel that some of your guesses are rather arbitrary. Just try to make as accurate a guess as possible with the information you get from the screen. I will give you no indication whether your guess was right or wrong.

The subsequent procedure used in the first session did not vary from that used for Groups I and II, with the exception that the *S* was not asked whether or not he had seen the stimulus.

In the second session, the *S*s matched their original responses to a list of items containing not only the 10 original stimuli, but also 10 additional items. Each of the additional items was a structural counterpart of one of the original stimuli, but was reversed in affective connotation. The original stimuli, with their counterparts indicated in parentheses, were as follows: barn (harm), seven (sexed), this (piss), line (lied), male sex organ (made six orders), vagina (vague), mother (matter), female sex organ (former sea origin), father (fatter), penis (pens).

Group IV. The thresholds for these *S*s were obtained in the same way as for Group III. The subsequent procedure differed in that the 10 stimuli were presented upside down and backwards on the screen, .3 log ft-L below the *S*'s previously determined thresholds. At this illumination level, the *S*s could clearly perceive the presence of the stimuli, but they did not detect the unusual orientation of the stimuli.

RESULTS AND DISCUSSION

Groups I and II

Because of similarity in experimental procedure, the results from Dixon's study and from Groups I and II may be compared directly. The thresholds for *S*s in Group I were determined by the procedure previously used by Dixon. Because of the possible unreliability of the thresholds obtained by Dixon's method, thresholds for *S*s in Group II were determined by a forced-choice method, and the stimuli were presented at a slightly higher illumination level than employed in Group I.

To determine if the *S*s could match their responses with the stimulus to which each response had been made originally, we followed Dixon in using a statistical test developed by Stevens (1939).

Four of the 16 *S*s in Group I made more than the chance number of correct matches, though none of the 4 was significant at the .05 level. One *S* obtained no correct matches, a result which is significant beyond the .02 level, in a direction opposite from that which was predicted. Dixon reports that 5 of his 7 *S*s made more correct matches than chance expectation, with 2 achieving significance beyond the .05 level. Two of Dixon's *S*s scored below chance, but neither score was statistically reliable.

The Stevens test permits the deviations from chance and the variances to be added over a number of *S*s to provide a significance test for the group's performance. Using this analysis for the data from Group I, a z value of -1.062 was obtained, which, for a one-tailed test, was not only insignificant ($p >$ .14), but in the opposite direction from Dixon's results. Application of a correction for discontinuity, suggested by Mosteller and Bush (1954), resulted in a z value of -1.128 ($p >$ .13). Applying the same analysis, without the correction for discontinuity, Dixon obtained a significant z value of 2.57 ($p < .005$).

While our Group I *S*s were unable to match meanings to the specific stimuli at better than chance level, the possibility existed that they might be able to make dichotomous discriminations between emotionally charged and neutral stimuli. To make this determination, a fourfold chi square was calculated for each *S* to determine the relationship between the affective classification of the stimulus matches (emotional or neutral) and the affective classification of the original stimulus. None of the 16 chi squares reached significance at the .05 level. Since three of the *S*s obtained values of chi square that reflected a relationship in the opposite direction, the over-all significance of the results could not be established by the usual method of adding independent chi square values. Instead, the Fisher-Pearson test was used (Gordon, Loveland, & Cureton, 1952). This test involves computing the exact probability for each obtained chi square value and then transforming them to $\log_e$. The sum of these transformed values is distributed as chi square with $2n$ *df* or, in the present case, 32 *df*. The application of this test to the data yielded a summed chi square value of 31.78, which clearly failed to attain significance ($p >$ .40).

Following Dixon's analysis, response latencies were examined to determine if shorter reaction times were given to emotionally neutral stimuli. Of the 16 *S*s in Group I, 7 manifested shorter average latencies for the neutral stimuli. In comparison, 5 of Dixon's 7 *S*s had shorter average latencies for the neutral items. Neither of these results was significant using the sign test. For Group I, a nonsignificant t value of .7902 (15 *df*, $p >$ .20) was obtained for the difference of the average response latencies to neutral versus emotional items.

To determine if sex differences were present, all the above analyses were repeated for the eight male and eight female *S*s, but no significant sex differences were revealed.

Next, Group II: Eight *S*s made more than the chance number of correct matches, two of which approached or exceeded significance ($p < .059$ and .015). On the other hand, three of the scores falling below chance had p values of .054, .048, and .017. Application of the Stevens test to the summed data from Group II resulted in a nonsignificant *z* value of .3354, corrected for continuity ($p > .35$).

Of the 16 chi square values relating the affective classification of the stimuli and the *S*s' stimulus matches, 8 were in the predicted direction, though none was significant at the .05 level. The Fisher-Pearson test yielded a chi square value of .2698 which, with 32 *df*, was not significant ($p > .50$). No significant differences were found for the reaction times of Group II to emotional or neutral stimuli, nor were there any significant differences attributable to sex.

The results from Groups I and II were consistent in failing to provide evidence that *S*s can perceive meaning at exposure levels too low for the *S*s to report the presence of the stimuli or at levels at which the *S*s can report the presence of the stimuli but cannot identify the stimuli.

Groups III and IV

While we were unable to confirm Dixon's findings that meanings can be perceived well below the recognition threshold, the possibility remained that meanings could be perceived prior to recognition, but where the gap between the presence-absence threshold and the recognition threshold was considerably less than used by Dixon. To test this possibility, Group III was employed.

Since the stimuli for this group were exposed at an illumination level at which 13% of the stimuli were correctly identified at the time of exposure, only those responses were analyzed in which the stimuli were incorrectly identified. In addition, 37 of the 128 responses to the stimulus phrases male sex organ and female sex organ were excluded because the *S* had correctly identified one or more of the words in these phrases.[2]

Application of the Stevens test to the reduced data from Group III revealed that 12 of the 16 *S*s made more correct matches than would be expected by chance.

Two of the *S*s' scores were significant beyond the .05 level, with one other significant beyond the .0001 level. None of the *S*s' scores which fell below chance expectation were statistically significant. Application of the Stevens test to the summed data from Group III yielded a *z* value of 3.61, corrected for continuity, which is significant beyond the .0002 level. This analysis revealed that when illumination levels were substantially higher than used in Groups I or II or by Dixon, *S*s were able to correctly match their initially incorrect responses with the stimuli which produced these responses. But, before attributing the better than chance performance of Group III to the unconscious perception of meaning, the results from Group IV must be examined.

It will be recalled that *S*s in Group IV responded to stimuli presented upside down and backwards at an illumination level only slightly lower than employed in Group III. Eight *S*s in Group IV obtained a number of correct matches which exceeded the chance expectation. Three of these scores were significant beyond the .05 level, while none of the scores falling below chance expectation were significant. The combined scores for Group IV, corrected for discontinuity, yielded a *z* value of 2.18, significant beyond the .02 level. The analysis revealed that even when the dimly illuminated stimuli were presented upside down and backwards, *S*s were able to match their incorrect identifications to a better than chance extent. This suggested that structural cues, e.g., differences in length among the stimulus words and phrases, must have been responsible for Group IV's performance, since the method of presentation, in all likelihood, precluded the perception of the meaning of the stimuli.

To determine if Group III *S*s were responding to the meaning of the stimuli in addition to structural cues, it was necessary to determine if Group III performed significantly better than Group IV. A nonsignificant *t* value of .51 was obtained (30 *df*, $p > .30$). Since the two groups did not differ significantly in their number of matches, it would appear that the performance of Group III is adequately explained in terms of nonmeaning cues such as word or phrase length. To check this conclusion and to rule out the implausible possibility that Group IV *S*s actually were perceiving the meaning of upside down and backward stimuli, another analysis was employed.

---

[2] If the *S* had perceived the word female in the phrase female sex organ, he had a cue in addition to meaning with which to improve the probability of his subsequently making a correct match. That initially correct identifications were highly determinate of the *S*s' subsequent matches is indicated by the finding that 95% of the correct identifications and 75% of the partial identifications led to correct matches.

It will be recalled that for Groups III and IV, the list of items to which the *S* was to match his responses contained not only the 10 stimuli actually presented, but also 10 additional items. Each of the additional items was a structurally similar counterpart of one of the original stimuli, but it was reversed in emotional connotation. If the *S* had originally responded to the meaning of the stimulus, his match could be expected to be to the actual stimulus to which the response had been made. On the other hand, if the *S* had responded to the structural characteristics of the stimulus, he would be expected to show an equal preference in his matches either for the correct stimulus or for its structurally similar counterpart. Therefore, for each *S*, we computed not only the number of matches to the correct stimulus, but also the number of matches to its structural counterpart. A difference score was obtained for each *S* by subtracting the number of matches made to structurally similar items from the number of matches made to the correct stimuli. The mean difference score of .125 obtained for Group III was positive in sign, but did not differ significantly from 0 ($t = .307$, 14 *df*, $p > .7$).

The above result offered no evidence that the *S*s were perceiving meaning in the absence of correct identification of the stimulus. There was a possibility, however, that the *S*s' response biases may have masked any effect attributable to the unconscious perception of meaning. The *S*s may have been reluctant to verbalize the emotional phrases and words as matches to their responses and, instead, chose the emotionally more neutral but structurally similar phrases. To evaluate the extent to which response bias may have operated in the data, we again compared the performance of Groups III and IV. For each *S* in each group, the difference score, i.e., the number of correct stimulus matches minus the number of matches to the structurally similar counterparts, were computed separately for the four neutral and six emotional stimuli. These scores then were analyzed by a modified three-way classification analysis of variance (groups, emotional versus neutral stimuli, and individuals).

TABLE 1

MEAN DIFFERENCE SCORES BY GROUPS AND AFFECTIVE CLASSIFICATION

	Emotional Stimuli	Neutral Stimuli
Group III	−1	1.125
Group IV	−1.063	.375

Only the main effect attributable to the affective classification of the stimuli was significant in the analysis ($F = 15.91$, 1 and 30 *df*, $p < .01$). The nature of these results can be seen in Table 1 which shows the mean of the difference scores by groups and affective classification. In both groups, the *S*s tended to choose the emotionally neutral member of the pair in making their matches. Thus, when an emotional stimulus had been responded to, the *S*s tended to match their response to the structurally similar, but emotionally neutral control. In the case in which a neutral stimulus had been present, the *S*s again tended to match their response to the neutral alternative, which in this case was the correct stimulus. It is clear from this analysis that the differential stimulus-matching of responses to emotional and neutral stimuli was attributable to the *S*s' bias to choose affectively neutral items for their matches.

The possibility may be raised that response biases masked the *S*s' ability to discriminate emotional items at the level of exposure used. This alternative is contradicted by the finding of no significant interaction of Groups III and IV with emotional or neutral words. As can be seen, the *S*s' response bias, while a significant effect, is not overriding (a mean difference of −1 for emotional stimuli, versus a possible difference of −6 if response bias was the sole determiner of choice). Thus, if perception of meaning was occurring, it should have mitigated the bias such that Group III would more accurately match responses made to emotional stimuli than would Group IV. However, as can be seen, the means are virtually identical offering no suggestion that perception of meaning was influencing these matches.

In addition, this analysis serves to render highly improbable the suggestion that Group IV *S*s were also perceiving meaning. Even if

Group IV *S*s were perceiving the meaning of the upside down and backward stimuli, it is to be expected that their perception of meaning should be less accurate than for Group III *S*s. Yet, Table 1 reveals this was not the case.

Our results are consistent in providing no evidence for the hypothesis that *S*s perceive the meaning of verbal stimuli at durations or illumination levels too low for verbal identification. Therefore, the results are quite at variance with Dixon's report. The question arises as to why the discrepancy between his results and the present findings.

First, the problem of sampling fluctuations must be considered. In Dixon's study, only seven *S*s were employed, of whom only five gave evidence of the unconscious perception of meaning. A second likely basis for the discrepancy between the two studies lies in the method of threshold determination employed by Dixon and the instructions given his *S*s. He used a modified method of limits and instructed his *S*s: "I am going to project words on the screen but they will be too dim for you to see" (Dixon, 1958, p. 30).

As has been amply demonstrated, threshold measurements depend importantly upon the instructions given the *S*s as well as the method employed to obtain the thresholds (Blackwell, 1953; Eriksen, 1958; Goldiamond, 1958). Dixon's instructions may have had the effect of giving the *S*s a set for not reporting the possibly perceived presence of the stimuli. The possibility exists that while Dixon's *S*s were unable to recognize the various words or even to perceive the letters, clearly, they nonetheless were able to pick up such structural cues as differences in word length or phrase length. These length cues were particularly marked in the set of stimuli employed in both Dixon's and the present study. That *S*s can utilize such cues is clearly demonstrated by the *S*s in our Group IV. For this group the verbal stimuli were presented upside down and backwards to eliminate all cues except differences in stimulus length. Yet, this group showed significantly better than chance matching.

Dixon fails to report whether the *S*s were sufficiently dark adapted before thresholds were obtained. It is possible that as dark adaptation progressed, Dixon's *S*s attained a degree of perceptual sensitivity not reflected in the thresholds obtained at the first of the experimental session. This, together with his rather inexact method of threshold measurement may account for the discrepancies between his results and ours.

SUMMARY

The hypothesis was investigated that the meaning of a verbal stimulus may be responded to without prior recognition of the stimulus. Four treatment groups were employed of 16 *S*s each. The experimental procedure used in the first two groups was very similar to that previously used by Dixon. The procedure consisted essentially of obtaining the *S*'s guesses to a series of verbal stimuli, dichotomized in affective connotation, which were presented at or slightly below the *S*'s previously determined absolute threshold. The results from these two groups were consistently at variance with results obtained by Dixon, for there was no evidence that the *S*s could make better than chance matches.

In the third treatment group, the same stimuli were presented at a higher level of illumination, such that 13% of the *S*s' guesses were correct identifications of the stimuli. The list of items to which the *S*s subsequently matched their guesses contained not only the 10 original stimuli, but also 10 additional items that were structural counterparts of the original stimuli, but reversed in affective connotation. A fourth treatment group made guesses to stimuli that were presented upside down and backwards to preclude the possibility that responses were made to the meaning of the stimuli. While both groups were able to make better than chance matches, the groups' performances did not significantly differ. Both groups were predominantly responding to structural differences among the stimuli, particularly to the markedly different length of the stimuli. It was concluded that no evidence was obtained that *S*s respond to the meaning of verbal stimuli prior to recognition of the stimuli.

REFERENCES

BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Bull. Engng. Res. Inst. U. Mich.*, 1953, **36**.

BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **51**, 24–29.

DIXON, N. F. Symbolic associations following subliminal stimulation. *Int. J. Psycho-Anal.*, 1956, **37**, 159–170.

DIXON, N. F. The effect of subliminal stimulation upon autonomic and verbal behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 29–36.

ERIKSEN, C. W. An experimental analysis of subception. *Amer. J. Psychol.*, 1956, **69**, 625–634. (a)

ERIKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80. (b)

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1958.

ERIKSEN, C. W., AZUMA, H., & HICKS, R. B. Verbal discrimination of pleasant and unpleasant stimuli prior to specific identification. *J. abnorm. soc. Psychol.*, 1959, **59**, 114–119.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–405.

GORDON, M. H., LOVELAND, E. H., & CURETON, E. E. An extended table of chi-square for two degrees of freedom, for use in combining probabilities from independent samples. *Psychometrika*, 1952, **17**, 311–316.

LAZARUS, R. S. Subception: Fact or artifact? A reply to Eriksen. *Psychol. Rev.*, 1956, **63**, 343–347.

LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

MOSTELLER, R., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *The handbook of social psychology.* Vol. I. Cambridge, Mass.: Addison-Wesley, 1954.

OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1957.

STEVENS, W. L. Tests of significance for extrasensory perception data. *Psychol. Rev.*, 1939, **46**, 142–150.

(Received September 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 440–447

THE CHILD'S DIFFERENTIAL PERCEPTION OF PARENTAL ATTRIBUTES[1]

JEROME KAGAN AND JUDITH LEMKIN

Fels Research Institute, Yellow Springs, Ohio

THE child's conceptualization of the parental roles and the specific characteristics he attributes to each parent are relevant to several theoretical issues. In describing the course of the child's identifications during the early school years, it has been suggested that the choice of parental models for identification is determined by the child's perception of parental differences in power, competence, and nurturance (Kagan, 1958; Maccoby, 1959). Moreover, Parsons and Bales (1955) have suggested that the mother's role is primarily concerned with maintaining warm, integrated, interpersonal relations. The paternal role prescription calls for instrumental skills and the ability to adapt to the environment. Thus, information on the child's perception of his parents may clarify problems associated with the identification process and test the validity of Parsons' dichotomy of parental roles.

Semantics is a second area which draws its theoretical statements from the individual's perception of the environment. Osgood, Suci, and Tannenbaum (1957) have reported that young adults use three conceptual dimensions in their categorization of a wide variety of objects, i.e., evaluative (good vs. bad), potency (strong vs. weak), and activity (active vs. passive). The content of the potency and activity dimensions resembles the stereotyped conception of male and female sex roles. For example, the adjectives large, strong, loud, heavy, brave, bass, rough, and rugged have the highest factor loadings on the potency factor. The adjectives agitated, sharp, ferocious, tense, hot, angular, active, and fast have the highest loadings on the activity factor. These adjectives are more commonly applied to males than to females, while their polar opposites are more characteristic of stereotyped female attributes.

Since the three factors discovered by Osgood et al. are relatively invariant over a variety of populations, it seems profitable to ask about the developmental history of this language learning. It is possible that these ingrained modes of conceptualization are a function, in part, of the child's perception of parental attributes. If a typical child perceives the father as stronger and more active than the mother, then one might expect the adult's conceptualization of "maleness" and "femaleness" to be colored by these earlier perceptions. Jenkins and Russell (1958) asked 540 college students to use the semantic differential in rating a variety of concepts, including mother and father, on 20 polar opposite adjectives. These young adults rated father, in relation to mother, higher on the dimensions cruel, successful, strong, and bad. It would be of interest to determine how early this pattern of labels is learned and whether the 6-year-old characterizes his father in a similar way.

This paper summarizes a study on the perception of parents by children 3 to 8 years of age. Several years ago, the senior author interviewed children 6 to 10 years of age and found that the father was perceived as more dominant and less friendly than the mother (Kagan, 1956). Gardner (1947) found similar results for children 10 to 13 years of age. Emmerich (1959) used a more indirect procedure with 4- and 5-year-olds and reported that children viewed the mother as more nurturant and less dominating and interfering than fathers.

In the Kagan study the questions were so phrased that the child was answering for himself and his parent, e.g., "Who punishes you, your mother or your father?" In the Emmerich study the questions were indirect and the child was responding in terms of "make believe" parents and children. No one has systematically studied the relationship between these two procedures. The present research had two purposes: to study the differential perception of mothers and fathers by young children with respect to power, punitiveness, competence, and nurturance; and to compare the protocols obtained by both direct and indirect methods of questioning.

---

[1] This research was supported, in part, by research grant (M-1260) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

METHOD

Subjects

The subjects were 32 boys and 35 girls ranging in age from 3–9 to 8–6 with a median age of 5–6 for the group. The children were enrolled in one of three nursery schools or the first grade of an elementary school at the time of the testing. Each child was tested individually at the school by the junior author (the experimenter). Several days prior to testing the experimenter visited the school in order to become acquainted with the subjects and reduce the effect of stranger anxiety on the test performance.

Procedure

Each child was asked four sets of questions in a standard order. The first set of questions involved color vs. form preferences in a simple concept formation task and is not relevant for the present report. Following this task the child was asked the remaining three sets of questions.

Set A. Perception of parents—indirect method. Initially, the experimenter showed the child a drawing of a family scene containing a father, mother, boy, and girl and said: "Here is a picture of a mommy, daddy, a boy, and a girl." She then replaced this family picture with two separate illustrations of a mother and a father and said: "Here are two pictures of just the mommy and daddy. Which one is the mommy? Which one is the daddy? I am going to ask you some questions about the mommy and daddy. You point to the one you think." The questions follow:

1. Who is the strongest one?
2. Who gives (the boy) (the girl) the most presents?
3. Who spanks the most?
4. Who is the boss at home?
5. Who does the (boy) (girl) like the best?
6. Who gets mad at the (boy) (girl)?
7. Who is the smartest?
8. If the (boy) (girl) did something bad and both mommy and daddy were home, who would punish the (boy) (girl)?
9. Who is nicer to the (boy) (girl)?
10. Who is the (boy) (girl) more scared of?
11. Who tells the (boy) (girl) what to do all the time?
12. Who kisses the (boy) (girl) the most?
13. Who would the (boy) (girl) like to grow up to be, who would the (boy) (girl) want to be like when (he) (she) grows up?

Set B. Perception of parents—indirect picture method. The experimenter then showed the subject a series of pictures (separate sets for boys and girls) in which a boy or a girl was in a situation that normally involved a parent figure but in which no adult was illustrated.[2] In some cases the hands of an adult were illustrated but no identification related to the sex of the adult was evident. The experimenter then said: "Here are some pictures which show a (boy) (girl) but the mommy and daddy are missing. You tell me who is missing—the mommy or the daddy." Descriptions of the pictures and the experimenter's explanatory statement to the child follow:

1. Some hands are shown giving the child an ice cream cone.

E: "Someone is giving the (boy) (girl) an ice cream cone—who is it? The mommy or the daddy?"

2. Some hands are shown reaching up for a toy that the child cannot reach.

E: "Someone is getting the (boy) (girl) the toy (he) (she) can't reach. Who is getting it?"

3. Some hands are shown spanking the child.

E: "Someone is spanking the (boy) (girl). Who is spanking? The mommy or the daddy?"

4. The child is held up high in the air by a pair of hands and has a smile on his face.

E: "Someone is picking the (boy) (girl) up in the air. Who is picking (him) (her) up? The mommy or the daddy?"

5. The child is cowering.

E: "Someone is yelling at the (boy) (girl). Who is yelling at the (boy) (girl)?"

6. Some hands are illustrated handing a gift wrapped box to the child.

E: "Someone is giving the (boy) (girl) a present. Who is giving (him) (her) a present?"

7. A dog is chasing a child who looks frightened and is running from the dog.

E: "The dog is chasing the (boy) (girl) and (he) (she) is running to someone for help. Who is the (boy) (girl) running to?"

8. The child is holding the hand of someone.

E: "Someone is taking a walk with the (boy) (girl). Who is it? The mommy or the daddy?"

9. The child is on a seesaw.

E: "Someone is on the seesaw with the (boy) (girl). Who is it? The mommy or the daddy?"

10. The child is sticking out his tongue with an angry expression.

E: "The (boy) (girl) is mad at someone. Who is (he) (she) mad at? The mommy or the daddy?"

Set C. Perception of parents—direct questioning method. After the pictures were administered the experimenter repeated the 13 questions that were asked in Set A, but the questions were asked in a different order. This time the experimenter phrased the question so that the child was talking about himself and his own parents. The experimenter said: "Now I am going to ask you a few questions about your mommy and your daddy." For example, the first question was phrased: "Who gives you the most presents? Your mommy or your daddy?"

The child's answers to all questions were recorded on a prepared form. If the child replied with neither mother nor father, an inquiry was made by the experimenter. If the child persisted in giving no answer or an answer other than mother or father the experimenter proceeded on to the next question. The majority of children understood the task and responded with either mother or father to all questions.

---

[2] Photostats of these pictures are available upon request.

TABLE 1
PERCENTAGE OF BOYS AND GIRLS ANSWERING MOTHER OR FATHER TO INDIRECT AND DIRECT QUESTIONS

Question	Indirect				Direct			
	Mother		Father		Mother		Father	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
1. Who is strongest	3.1	2.9	96.9	97.1	3.1	5.8	93.8	94.2
2. Who gives presents	78.1	57.1	15.7	37.1	60.6	45.7	27.0	20.0
3. Who spanks most	37.5	28.5	59.4	65.7	27.0	37.1	53.1	51.3
4. Who is boss	9.3	25.7	87.6	71.4	15.7	14.3	56.2	68.6
5. Who child likes	47.1	48.7	47.1	40.0	31.4	62.8	40.6	22.8
6. Who gets mad	40.6	25.7	53.1	71.4	47.1	40.0	50.2	51.3
7. Who is smartest	9.3	37.1	84.5	62.8	18.6	28.5	75.2	59.8
8. Who punishes	37.5	25.7	59.4	71.4	27.0	25.7	56.2	45.7
9. Who is nicer	60.6	68.6	37.5	31.4	56.2	59.8	31.4	31.4
10. Who child is scared of	31.2	11.4	65.6	85.6	31.2	20.0	31.2	57.1
11. Who tells child what to do	43.7	54.3	50.2	45.7	40.6	54.3	40.6	37.1
12. Who kisses child most	78.1	42.9	21.9	54.3	75.2	51.3	15.7	34.3
13. Who will child grow up to be	3.1	88.1	90.6	5.7	3.1	82.8	90.6	2.9

RESULTS

Differential Perception of Mother and Father

Table 1 presents the percentage of the boys and girls responding mother or father to the 13 questions in Set A (indirect method) and Set C (direct method).

Table 2 presents the percentage of boys and girls responding mother or father to the 10 pictures shown in Set B.

Initially, both boys and girls were divided into young vs. old in order to test for age differences on each of the three sets of questions. Since there were no statistically significant differences between the answers of the young and old children to any of the questions, the results presented are for all boys and all girls.

The data in Table 1 reveal that, except for Questions 5, 10, and 12, the answers to the indirect questions were quite similar to those obtained with the direct questions. For both the direct and indirect methods boys and girls saw the father, in relation to the mother, as stronger, the boss of the house, smarter, and the major agent of punishment (Questions 1, 3, 4, 7, and 8). To the question, "Who is the child more scared of?" boys and girls answered father more than mother on the indirect query. All of these differences were significant at the .05 level and all probability values presented are for two-tailed tests unless otherwise stated. The difference between mother and father answers to the direct question, "Who are you more scared of?" was not statistically significant for boys but was significant for girls ($p < .05$). That is, girls labeled father as more fear provoking than mother on both kinds of questions. To the indirect question, "Who gets mad?" the girls replied father more often than mother ($p < .05$). There was no difference for boys between father and mother answers on either type of question. Thus, girls labeled the father as more fear provoking and prone to anger than did the boys.

The mother, in relation to the father, was viewed as the giver of presents and nicer (Questions 2 and 9) on both methods ($p <$ .05). To the question, "Who kisses the most?"

TABLE 2
PERCENTAGE OF BOYS AND GIRLS RESPONDING MOTHER OR FATHER TO THE TEN PICTURES

Picture	Boys		Girls	
	Mother	Father	Mother	Father
1. Child given ice cream	59.4	40.6	82.7	17.3
2. Hands reaching for toy	25.0	75.0	11.4	88.6
3. Child being spanked	40.6	56.3	31.4	68.6
4. Child picked up in air	21.9	78.1	14.3	85.7
5. Child cowering	12.5	87.5	22.8	77.2
6. Child given present	87.5	12.5	68.6	31.4
7. Child running from dog	28.1	71.9	31.4	68.6
8. Child taking walk	53.1	46.9	51.4	45.6
9. Child on seesaw	31.2	50.0	37.1	28.5
10. Child mad	50.0	46.9	28.5	62.5

boys replied mother more often than father to both methods ($p < .05$). This difference did not hold for girls. The question, "Who tells the child what to do?" yielded no significant difference between the frequency of the mother and father answers.

The responses to the 10 pictures were consistent with these results. To Pictures 1 and 6, where a child was receiving a desirable object, both boys and girls responded mother more than father ($p < .05$). The pictures illustrating punishment (3 and 5) elicited more father responses ($p < .05$). The pictures depicting a child in a situation where he needs or receives assistance, strength, or protection (Pictures 2, 4, and 7) also elicited more father than mother answers ($p < .05$). There were no significant sex differences for any of the 10 pictures taken separately. The two pictures that were more neutral in content (8 and 9) did not produce significant differences between frequency of mother or father answers. Finally, it is of interest to note that to Picture 10, where the child was described as mad at someone, boys showed no significant difference between mother and father answers while girls responded father more than mother ($p < .05$). This result agrees with the answers to the indirect question: "Who gets mad at the child?" The girls replied father more often than mother while there was no difference in father-mother answers for the boys. It appears that with indirect procedures girls are more likely than boys to label the father as potentially hostile.

The results for all three methods agree with the research summarized earlier and reaffirm the statement that young children see mothers as more nurturant but less punitive, less fear arousing, and less competent than fathers.

Method Differences

There were only three questions that produced slightly different answer patterns in comparing the indirect with the direct method of questioning (Questions 5, 10, and 12). None of these differences were statistically significant for both sexes. To the question, "Who is the child more scared of?" the boys were more likely to answer father to the indirect than to the direct question ($p < .05$). To the question, "Who kisses most?" the girls answered father more to the indirect than to the direct question ($p < .05$). Thus, when the question dealt with fear of father boys showed the greatest shift in answers between the indirect and direct procedure. When the question dealt with physical affection girls showed the greatest shift between the two types of questions. To the question, "Who does the child like the best?" girls showed a small but nonsignificant increase in mother answers to the directly phrased question.

It is suggested that perhaps these three questions were more anxiety arousing than the remaining 10 and that the direct questioning produced sufficient anxiety to elicit some defensive processes. The direct question, "Who kisses you most?" may have elicited sexually tinged anxiety because of the implications of a physically affectionate relationship between the girl and her father. The fact that this shift did not occur for boys suggests that recognition or report of physical affection between a child and the opposite sex parent may be less anxiety arousing for boys than for girls.

The possibility that the direct method was more anxiety arousing finds support in the fact that the direct method elicited more evasive answers than the indirect method. Although the instructions emphasized that the child respond either mother or father, some children gave no response, answered both mother and father, or named another family figure like brother or sister. These "other" answers were much more frequent to the direct questions than to the indirect ones. The average number of children giving one or more "other" answers to the 13 indirect questions was 2.4 while the average was 10.4 for the direct questions ($p < .01$). About 15% of the children gave at least one "other" answer to the direct questions while only 3% of the group did so to the indirectly phrased questions. There were no sex differences in number of evasive replies.

Sex Differences

Sex differences did occur on some of the questions. The most striking sex difference occurred to the question: "Who would you like to grow up to be?" All children answered with the same sex parent on both indirect and direct methods ($p < .001$). The questions that dealt with affectionate and nurturant relations between child and parent also produced sex differences. To the indirect question, "Who kisses most?" girls chose father and boys chose mother ($p < .01$). This difference did not hold for the direct question and it has been suggested that this might be due to the girl's anxiety over responding father to the direct query. To the direct question, "Who does the child like the best?" boys and girls tended to choose the same sex parent over the opposite sex parent ($p < .05$). Finally, the indirect question, "Who gives the most presents?" produced the slight tendency for girls to respond father more than boys ($p < .08$). The girls showed a tendency, on the indirect questions (Questions 2 and 12), to label the father as more nurturant than boys. However, the girls also showed a slight tendency on the indirect set to see the father as more punitive, more prone to anger, and more fear arousing than boys.

Sex and Method Differences for Content Clusters

Boys and girls differed in the degree of punitiveness and affection assigned to the father, despite the fact that all children agreed that mother was more nurturant and less punitive than father. In order to test the significance of this trend those questions that dealt with similar contents were considered as a cluster. The three clusters analyzed were punitiveness-hostility (Questions 3, 6, 8, and 10), nurturance-affection (Questions 2, 9, and 12), and competence-power (Questions 1, 4, and 7). The interitem correlations among the indirect questions in each of the three clusters were all positive and justified the pooling of these items. A child's score for father or mother answers on the punitiveness-hostility cluster ranged from 0 to 4; for the nurturance-affection and competence-power clusters the range was 0 to 3. The distributions of these scores for father and mother answers for each set of questions were highly skewed. Each distribution was divided at the median and chi square tests (with Yates' correction) were applied to assess sex and method differences for each of the three clusters. Table 3 presents the results of this analysis.

Sex differences. The direct method produced no significant sex differences. On the indirect method, the girls, in comparison to boys, characterized father as more punitive, more nurturant, but less competent ($p < .10$, .01, .01, respectively). The boys ascribed more competence to father, while girls attributed more competence to mother. This sex difference did not appear on the direct questions, suggesting that the indirect questions may be reflecting the child's wishes. Since there should be minimal anxiety over ascription of intelligence, strength, and power to a parent the fact that boys and girls show a decrease in same sex parent answers to the direct questions raises the possibility that the indirect answers were influenced by the child's desire to view the same sex parent as the more competent one.

Method differences. Boys were more consistent than girls across the two methods. To the direct questions, girls ascribed less punitiveness and less affection to father than they did on the indirect set. The boys' answers on

TABLE 3

SEX AND METHOD DIFFERENCES FOR THE PUNITIVE, NURTURANT, AND COMPETENCE CLUSTERS

Cluster	Parent	Sex Differences		Method Differences	
		Indirect	Direct	Boys	Girls
Punitive	Father	G > B*	*ns*[a]	*ns*	I > D***
	Mother	*ns*	*ns*	*ns*	*ns*
Nurturant	Father	G > B***	*ns*	*ns*	I > D**
	Mother	*ns*	*ns*	*ns*	*ns*
Competent	Father	B > G***	*ns*	I > D***	*ns*
	Mother	G > B***	*ns*	*ns*	I > D*

Note.—B = Boys, G = Girls, D = Direct method, I = Indirect method.

[a] *ns* = not significant; $p > .15$; two tails.

* $p < .10$; two-tailed test.

** $p < .05$; two-tailed test.

*** $p < .01$; two-tailed test.

these two clusters were not significantly different across the two methods. The phi coefficients evaluating the reliability of father answers across the two methods were significant for boys at the .05 level for all three clusters (phi = .38, .39, and .29 for the punitive, nurturant, and competence dimensions, respectively). For the girls, the corresponding coefficients were .17, .00, and .41, the latter value significant at the .05 level.

Punitive and nurturant clusters of items were also selected from the 10 pictures in Set B. These answers were correlated with the responses to the punitive and nurturant clusters on the indirect and direct questions. Pictures 1, 4, and 6 were used for the nurturant cluster and Pictures 3, 5, and 12 for the punitive-hostile cluster. The relationship between answers to pictures and questions was assessed by phi coefficients. The answers to the indirect, punitive, and nurturant questions were significantly correlated with the answers to the punitive and nurturant pictures. That is, those children who labeled the father as punitive-hostile on the indirect questions also labeled the father as punitive-hostile on Picture Cluster 3, 5, and 12 (phi = .38; $p < .01$). Those children who labeled the mother as nurturant on the indirect question cluster replied mother to Picture Cluster 1, 4, and 6 (phi = .33; $p < .01$). These results held for both sexes. The answers to the direct, punitive, and nurturant clusters were not significantly related to the answers on the punitive and nurturant pictures.

DISCUSSION

These results are in remarkably strong agreement with those of other studies (Emmerich, 1959; Finch, 1955; Gardner, 1947; Kagan, 1956) and indicate that children from preschool to preadolescence label the mother as more nurturant than the father. Father is labeled as more punitive, more competent, and a more frequent source of power and fear. Although all children agree that father is punitive and mother nurturant, the girls view the father with more ambivalence than the boys. That is, the girls label the father as both more hostile and more affectionate than the boys. It is suggested that the 6-year-old boy has begun to identify with his father, and is forced to repress some of the father's fear arousing and undesirable traits. The young boy, therefore, conceptualizes the father in a warmer and more benign light. The girl sees the father both as a more distant figure (because of the lack of identification with him) and as a potential love object and source of affection. This *post hoc* explanation would account for the girls' ambivalent, conscious perception of the father.

The data support Parsons' hypothesis that the mother role is predominantly nurturant while the paternal role is characterized by instrumental skills and competence. The results also agree with Brim's (1957) suggestion that each parent forms a different type of social relationship with a son or a daughter, e.g., the father-daughter relation is a different social system than the father-son interaction.

The authors do not believe that the child's perception of the parents is solely a function of each parent's behavior toward the child. The mass media, peer communications, and the content of children's books, all paint the adult male as competent, aggressive, and powerful. These sources of communication describe the mother as loving, weak, and subordinate (Bennett & Cohen, 1959; Child, Potter, & Levine, 1946). It is likely that the child's answers to our questions were influenced by these stereotyped conceptions of the mother and father roles. At a more speculative level, it is suggested that these extrafamilial communications also influence the child's perception of his parents.

That is, if a 5-year-old views men as potentially more aggressive and hostile than women there is reason to believe that he will generalize this expectation to his own father and view the latter as potentially more aggressive than he may be in reality. The child's conceptualizations of the significant roles and people in his life are apt to be the result of an interaction between his direct experiences with these social entities and the stereotyped labels he acquires through contact with symbolic representations of these roles.

The acquisition of meaning and perception of parents. The differential perceptions of father and mother seem congruent with the meaning factors described by Osgood et al. (1957). The reader will recall that Osgood found that the adjectives large, strong, rough, hard, and heavy, which suggest power and masculinity, were positively intercorrelated. Each of these words showed zero or negative correlations with adjectives like kind, pleasant, and nice, which imply nurturance and femininity. It is suggested that the latter group of adjectives is more congruent with the child's conception of mother while the former is more characteristic of the concept of father. The present data suggest that the conceptual associations of adults, with respect to the attributes strength, power, and nurturance, have historical roots in the child's perception of his parents. Osgood reports that the word "nice" is closer in the semantic space to femaleness than it is to maleness. In the present study, the question, "Who is nicer to the child?" elicited more mother than father answers from both boys and girls. Thus, the purely conceptual association between the words "nice" and "female" discovered in college students is present in the conceptualization of mother and father by preschool children.

As mentioned earlier, Jenkins and Russell (1958) found that college students rated father, in relation to mother, as higher on the traits cruel, successful, strong, and bad. The current sample of children differentially labeled their parents in a similar way. This similarity suggests that such language learning begins very early in the child's conceptual development.

Identification processes. In an earlier paper on identification (Kagan, 1958) it was stated that the model the child would most likely choose for identification was the one who was viewed as competent and in control of sources of gratification. The present data suggest that girls view the father as a more competent and powerful figure than the mother. However, to the question, "Who do you want to be like when you grow up?" the girls answered mother. Thus, the role model that the girls have chosen is not the one whom they view as the most competent. It is tempting to speculate on the implications of this finding. Since the girl's identification influences both her conception of herself and her behavioral tendencies with other people, it is possible that the passivity of women with men is based, in part, on the child's belief that the father is stronger and more competent than the mother. These speculations are in essential agreement with other research on the sexual preferences of children. Brown (1956) asked 5- and 6-year-old children to select preferred activities and role behaviors. Although children typically preferred their own sex role, twice as many girls as boys indicated a preference for both male and female activities and roles. Moreover, boys showed a significantly greater preference for their own sex role than girls showed for the female role. On choosing between wanting to be a boy or a girl, only 10% of the boys said they desired to be a girl, while about a third of the girls preferred the male role. Smith (1939) has reported that, with age, both sexes come to have a better opinion of boys and progressively poorer opinion of girls during the ages 8 to 15. Smith asked 200 children whether boys or girls possessed a greater degree of desirable vs. undesirable traits. By age 14, girls felt boys were as high on the desirable attributes as were the girls.

The data suggest that girls have a more anxiety arousing identification and a less positive self-image than boys and that this phenomenon is influenced by the differential perception of fathers and mothers during early childhood.

Methodological considerations. This investigation suggests two tentative statements about the differential validity of indirect and direct methods of questioning with young children. Although there was a moderate degree of consistency across both procedures, the direct method elicited more evasive answers. The greatest differences between the two procedures occurred on those questions that might be regarded as anxiety arousing. One of the advantages of using both direct and indirect methods in a diagnostic or research setting is to obtain clues as to sources of anxiety. Lack of consistency between the answers to the two methods may indicate some anxiety associated with the content of the question.

Since the indirect questions elicited less evasiveness one might conclude that this technique gave a more accurate appraisal of the child's perceptions. The responses to the competence and power questions tend to contraindicate this conclusion. The children were more likely to see the same sex parent as smart and the boss on the indirect than on the direct questions. Since these questions should not elicit high anxiety it is possible that the child's wish to view the same sex parent as more competent influenced his answers. Each method has unique characteristics and the combined use of the methods in research and diagnosis is perhaps the most sensible strategy.

SUMMARY

A group of 67 children were individually interviewed with three different methods to obtain information on perception of the parents with respect to nurturance, punitiveness, source of fear, and competence.

Indirect questions: The children were first shown line drawings of a mother, father, and child and asked 13 questions about this "make believe" family. The questions pertained to parental nurturance, competence, punitiveness, and source of fear: e.g., "Who gives the child the most presents?" "Who is the child more scared of?"

Picture method: The children were then shown 10 different pictures illustrating a child in situations that usually involved a parent, but the parent was not illustrated. The child was asked to name who was missing from the picture, the mother or the father.

Direct questions: Finally, the child was asked the same 13 questions asked in the indirect method (in a different order) but the questions were phrased in terms of the child and his own parents: e.g., "Who gives you the most presents, your mother or your father?" "Who are you more scared of, your mother or your father?"

The results suggested the following conclusions:

On all three methods the fathers were seen as more fear arousing, more competent, and more punitive than mothers. Mothers were seen as "nicer" and more likely to give presents. This result held for both sexes.

There were a few sex differences. Boys and girls chose the same sex parent as the model they wished to emulate and the parent they "liked the best." They chose the opposite sex parent as the one who "kissed the most." Although all children labeled the father as punitive and mother as nurturant, the girls, in comparison with the boys, labeled the father as both more punitive and more affectionate. The girls seemed to perceive the father with more ambivalence than the boys.

The direct questions elicited significantly more evasive answers than the indirect questions. It was suggested that the indirect questions were most likely to elicit material different from direct questioning when the content of the question was anxiety arousing or was related to a wish of the child.

Girls stated a desire to be like the mother but perceived the father as wiser and stronger. It was suggested that girls have an anxiety arousing identification for their identification model is perceived as the least competent of the two parents.

The results supported both previous research in this area and Parsons' hypothesis regarding the differentiation of parental roles. The developmental implication of these results for Osgood's meaning factors was discussed.

REFERENCES

BENNETT, E. M., & COHEN, L. R. Men and women: Personality patterns and contrasts. *Genet. psychol. Monogr.*, 1959, **60**, 101–153.

BRIM, O. G., JR. The parent-child relation as a social system: I. Parent and child roles. *Child Develpm.*, 1957, **28**, 343–364.

BROWN, D. G. Sex-role preference in young children. *Psychol. Monogr.*, 1956, **70**(14, Whole No. 421).

CHILD, I. L., POTTER, E. H., & LEVINE, ESTELLE, M. Children's textbooks and personality development: An exploration in the social psychology of education. *Psychol. Monogr.*, 1946, **60**(3, Whole No. 279).

EMMERICH, W. Young children's discriminations of parent and child roles. *Child Develpm.*, 1959, **30**, 404–420.

FINCH, HELEN M. Young children's concept of parent roles. *J. home Econ.*, 1955, **47**, 99–103.

GARDNER, PEARL L. An analysis of children's attitudes toward fathers. *J. genet. Psychol.*, 1947, **70**, 3–28.

JENKINS, J. J., & RUSSELL, W. A. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

KAGAN, J. The child's perception of the parent. *J. abnorm. soc. Psychol.*, 1956, **53**, 257–258.

KAGAN, J. The concept of identification. *Psychol. Rev.*, 1958, **65**, 296–305.

MACCOBY, ELEANOR E. Role taking in childhood and its consequences for social learning. *Child Develpm.*, 1959, **30**, 239–252.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PARSONS, T., & BALES, R. F. *Family, socialization and interaction process.* Glencoe, Illinois: Free Press, 1955.

SMITH, S. Age and sex differences in children's opinion concerning sex differences. *J. genet. Psychol.*, 1939, **54**, 17–25.

(Received September 28, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 448–456

EFFECTS OF WHITE AUTHORITARIANISM IN BIRACIAL WORK GROUPS[1]

IRWIN KATZ AND LAWRENCE BENJAMIN
New York University

ALTHOUGH the California F Scale was originally developed to study the prejudiced personality, there have as yet been no investigations of the relationship between F Scale scores and actual behavior toward minority group members. It is not known, for example, how the social behavior of authoritarian whites differs from that of nonauthoritarians in various types of face-to-face contact with Negroes. Nor is it known how Negroes react to whites who differ in this regard. The present experiment attempted to study social interaction in work groups composed of two Negro men, and two white men with either high or low F Scale scores, under various conditions of reward and prestige.

Two lines of previous research on the F Scale are pertinent to this problem. First, it has been established that authoritarians are strongly prejudiced against cultural minorities. Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) reported positive correlations between the E (Ethnocentrism) and F Scales. Subsequent research, according to Christie and Cook (1958), supports this demonstrated relationship between the F Scale and prejudice. Second, Thibaut and Riecken (1955) found that authoritarians were more openly hostile against persons of low status than of high status, while variations in the status of the instigating person produced little difference in the expression of hostility by nonauthoritarians.

Thus, in the absence of external restraints one would expect authoritarian whites, as compared with nonauthoritarians, to aggress more against members of a low-status cultural minority, e.g., Negroes. But how do authoritarians behave toward Negroes when openly hostile acts are likely to bring forth retaliation, or punishment by an authority? Contact situations in which whites and Negroes have roughly equal status and in which there exists an official policy of nondiscrimination are becoming increasingly common in American life, both military and civilian. The present study was designed to observe interaction in a biracial situation of this kind. All men had identical job roles, received identical pay, and were treated alike in all respects by a white supervisor.

In addition to authoritarianism we were interested in two variables that gave promise of interacting with it, with both practical and theoretical implications: type of reward (group vs. individual) and level of group prestige.

Reward. Under a group bonus system, as compared with individual, noncompetitive bonuses, Negroes would have greater ability to retaliate against white unfriendliness, since by not cooperating on tasks they could prevent the latter from earning extra money. On the other hand, if group efforts were successful white men would tend to see Negro group members as agents of satisfaction, and consequently would be inclined to react more favorably to them. The first factor would inhibit anti-Negro behavior and the second would reduce instigations to such behavior. On the assumption that authoritarians are basically more hostile toward Negroes than are nonauthoritarians these two factors might be expected to produce greater differences in the behavior of authoritarian subjects (*S*s) than of nonauthoritarian *S*s.

Prestige. It is a reasonable assumption that in a condition of high group prestige, as compared with a neutral condition, white *S*s would tend to see Negro *S*s more as fellow members of a small élite and less as representatives of a disliked cultural minority, with the result that their behavior toward Negro *S*s would be more acceptant. On the supposition that authoritarians react to people primarily on the basis of perceived status, we expected that their social behavior would be more strongly affected by variations in group prestige than would the social behavior of nonauthoritarians.

---

[1] This is the second report on research carried out under Contract Nonr-285(24) between the Office of Naval Research and New York University. The project is under the direction of Irwin Katz. Judith March was a behavioral observer in the present experiment, in addition to the junior author.

Racial differences for all conditions combined. In an experiment on biracial groups Katz, Goldston, and Benjamin (1958) found that white men were more active, communicative, and productive than their Negro companions, and tended to talk mainly to one another, while Negroes addressed most of their remarks to whites and largely ignored one another. In the present study we wished to ascertain whether similar differences would occur on a wider range of measures when Negro and white group members were more closely matched on intelligence, and were repeatedly made aware of their mental equality, even to the point of being given tasks which were rigged to insure similarity of performance.

METHOD

Negro and white men college students were employed to work in groups of four—two Negroes and two whites—on a series of tasks that were described as "materials that are being developed for vocational aptitude tests." Each group worked a total of 12½ hours, divided into four or five sessions, under a particular combination of white authoritarianism or white nonauthoritarianism, high prestige or neutral prestige, and group reward or individual reward. A 2 × 2 × 2 factorial design was used, with two groups in each cell. As dependent variables, social interaction, social perception, social influence, and productivity were measured by means of observers' classifications of ongoing behavior, questionnaires, and objective measures of output.

Selection of Ss and matching on intelligence. All *S*s were from colleges and universities in New York City and between the ages of 19 and 30. Fifty-six Negroes and 160 whites were given the Verbal Reasoning and the Language Usage Tests of the Differential Aptitude Tests. The Verbal Reasoning Test became the primary instrument for matching men on intelligence, with scores on Language Usage being used to decide borderline cases.[2] The 32 highest scoring Negroes were selected from the original Negro pool of 56 men and matched as closely as possible with 32 whites (of whom 16 were identified as authoritarians and 16 as nonauthoritarians, by a method to be described below). Then, Negro-white matched pairs were randomly assigned to experimental groups. The mean Verbal Reasoning score of each race was 33, which corresponds to Percentile 65 in a national sample of high school seniors (Bennett, Seashore, & Wesman, 1947).

In the Negro sample the age range was 19 to 30 and the mean was 23.3. With one exception all Negro *S*s had grown up in the North. In the white sample the age range was 19 to 28 and the mean was 21.2. Twenty-two white *S*s were Jewish, five were Catholic, and five were Protestant.

Establishment of the experimental conditions. The laboratory consisted of a behavior room and an adjacent observational room connected by a one-way window and a one-way sound system. All instructions were given orally by the experimenter (*E*). Each group was told that it was to try out various potential test materials. *S*s were also informed that they had been placed together because of the similarity of their DAT scores.

At the time of taking the DAT all men were administered a slightly modified version of the Webster, Sanford, and Freedman F Scale (1955).[3] Those in about the top third of the total white distribution of 160 scores were considered as high scorers and those in about the bottom third as low scorers. Sixteen high scorers and 16 low scorers were matched with Negro *S*s on Verbal Reasoning scores and then assigned to work groups so that eight groups had 2 authoritarian whites and eight groups had 2 nonauthoritarian whites. The mean Verbal Reasoning scores of both Negro and white *S*s in authoritarian groups and in nonauthoritarian groups were equated.[4]

All men were told that a bonus of 35¢ an hour, in addition to the base pay of $1.50 an hour, would be awarded for "effort." For induction of group reward *S*s in eight groups were told that the bonus would be based on the effort of the group as a whole. For individual reward *S*s in eight groups were told that the bonus would be awarded to each man individually on the basis of his own effort. It was stressed that the individual bonus was noncompetitive.

In eight high prestige groups men were told at the first session that on the basis of their DAT scores they were expected to perform better than other groups. At the start of the second and all subsequent sessions high prestige groups were shown graphs which favorably compared their scores on various tasks with those of 15 fictitious groups. Eight neutral prestige groups received no information about their expected performance or actual performance relative to other groups.

---

[2] It might be objected that the DAT was designed for use with high school samples, and would have proved too easy for college students. However, this was not the case for our Negro *S*s with whom white *S*s were matched.

[3] We omitted 9 inappropriate items (all those under *D.3* "preference for traditional feminine role") from this 149-item scale that had been developed on a sample of Vassar undergraduates. It was used in preference to the original California scale because of its greater length and greater concentration on personality as opposed to ideology. In a sample of 100 white New York University men undergraduates, the Vassar scale showed a correlation of .79 with the California scale. In another sample of 67 from the same population its correlation with the Christie, Havel, and Seidenberg version of the F Scale (1958) was .69, and .26 with these authors' "acquiescence score." The latter correlation indicates that acquiescent response set was a factor of only minor importance in the Vassar F Scale scores of NYU men.

[4] For the total white distribution of 160 scores on the modified Vassar scale, $M = 540$ and $SD = 65$. Cutoff points for high scorers and low scorers were 580 and 497, respectively. For the total Negro pool of 56 men, $M = 539$ and $SD = 69$. Mean Negro F Scale scores for different experimental conditions ranged from 532 to 540.

The tasks. Ss worked on two 50-minute construction tasks with Erector Set materials, in which all had to build identical models from a picture. They were permitted to help one another.

Two human relations problems were discussed, each for 30 minutes. Groups had to agree unanimously about suitable answers to a set of questions. Before each discussion Ss answered the same questions privately.

A maze problem was rigged so that the average scores of Negroes and whites in the same group would be the same. The apparatus consisted of a 3 × 3 foot vertical wooden panel covered with bolt heads spaced 1 inch apart, and a metal stylus. The board was wired electrically so that when a "correct" bolt head was touched with the stylus a bulb lighted. A vertical path had to be learned. Ss worked individually, taking one trial at a time in rotating order. The number of trials required for an errorless run was secretly controlled from the observation room.

The maze apparatus contained two more complex paths which the men worked on as a team. Before each move Ss registered their opinions privately, after which they were required to reach a unanimous group decision through discussion. The latter determined the actual move.

A set of 20 multiple-choice abstract reasoning problems was rigged so that the average scores of Negroes and whites in a given group would be the same. After each problem the men announced their answers, and then the "correct" answer was announced by *E*.

Every group worked as a team for a maximum of 40 minutes on the Hanfmann-Kasanin Concept Formation Test.

Pairs of men worked on a two-man version of the ball-and-spiral game that was used by French (1953) and others. An open spiral tract $2\frac{1}{2}$ inches wide was mounted on a 25 × 25 inch base, at each corner of which was a handle. Two men faced one another, each grasping two handles, and tried to maneuver a rubber ball up the tract and into a hole at the top. Every *S* worked $\frac{1}{2}$ hour with the group member of his own race and $\frac{1}{2}$ hour with a randomly selected group member of the other race. The order of pairing was counterbalanced for the various groups.

The order of tasks was as follows: Individual Maze, Construction 1, Group Maze 1, Abstract Reasoning, Human Relations 1, Construction 2, Hanfmann-Kasanin, Human Relations 2, Group Maze 2, Spiral.

Dependent variables. Communications were classified as they occurred by one of two observers, who were ignorant of the experimental treatments accorded particular groups. Twelve categories were used, as presented in Table 1 below. (Some were used on more than one task.) Interobserver correlations (rho) on number of communications initiated in various groups were computed separately for Negroes and whites. They ranged from .72 to .96 for different categories, with a median of .84. Six categories yielded, in addition to frequency scores, measures of the social direction of communications. For this purpose we employed an "S score" (Katz et al., 1958), which has to do with the proportion of remarks directed by an individual to own-race and to other-race recipients. On these scores the range of interobserver correlations was .65 to .92, with a median of .81.

Measures of opinion change under group influence were obtained on the group mazes and the human relations problems.

A terminal questionnaire included items on social preference, personality ratings, and evaluations of performance. Satisfaction with the group experience was measured with a 15-item Likert-type scale developed by Gibb (1954). Defensiveness toward the group was measured with a 25-item Likert-type scale reported by Smith (1956).

RESULTS AND DISCUSSION

Main Effects of Authoritarianism on Communication

In evaluating the effects of the independent variables on communication, the various categories of behavior were treated separately for each type of task, yielding a total of 21 measures of amount of communication for each race.[5] Each set of behavior scores was submitted to analysis of variance, taking into account three major variables—authoritarianism, reward, and prestige. We will report the main effects of authoritarianism only, but all interactions.[6]

White. Table 1 shows *F* ratios for main effects of the authoritarianism variable on number of communications initiated by white Ss, as well as interactions of authoritarianism with reward and prestige. Four main effects had probabilities below .10, including two significant main effects ($p < .05$) without interactions. Both of the latter occurred on spiral-mixed pairs: authoritarian whites accepted more action proposals from their Negro partners and rejected fewer than did nonauthoritarians. It is not surprising that the clearest behavioral differences between the two types of white Ss occurred when they were paired with Negroes on the spiral, for this task was intended to maximize interdependence by requiring close motor coordination between partners.

Negro. Table 2 shows all *F* values for main

---

[5] This does not include the group mazes, which were omitted because of a procedural error that resulted in time samples of indeterminate length.

[6] Similar analyses were performed on scores indicating direction of communication (S scores), but the effects of the experiment variables were so few and inconsistent that they are not included in this report.

TABLE 1

VALUES OF F FOR MAIN EFFECTS OF AUTHORITARIANISM AND INTERACTIONS WITH PRESTIGE AND REWARD ON NUMBER OF COMMUNICATIONS INITIATED BY WHITES

Task and Category	Main Effects		Interaction		
	Author. Higher	Nonauthor. Higher	A/P	A/R	A/P/R
Construction					
Asks for help	5.73**		10.52**	.68	12.60***
Gives help	.45		.40	.52	.10
Orients group toward goal		1.20	1.19	1.01	9.57**
Human Relations					
Disagrees	2.72		.97	.64	.39
Agrees	.90		.53	.48	17.10***
Gives information or opinion	.45		1.01	.66	1.20
Asks for info., etc.	.79		.89	.54	1.42
Hanfmann-Kasanin					
Disagrees	.37		.11	.85	.15
Asks for opinion		.32	.46	.28	.48
Gives opinion		1.05	.32	.43	.53
Orients group		1.24	.79	.67	.28
Spiral—Mixed Pairs					
Accepts proposal	16.15***		.61	.47	.55
Proposes action	.88		8.44**	.22	1.66
Praises other		.25	7.50**	.19	.50
Expresses hostility		.36	1.62	.59	.73
Rejects proposal		5.86**	.40	.14	.76
Spiral—White Pairs					
Expresses hostility	3.76*		8.43**	.93	1.16
Proposes action	1.08		5.73**	.88	.64
Rejects proposal	.90		.34	.28	.78
Accepts proposal		1.44	.17	1.90	3.32
Praises other		1.65	12.40**	4.31*	.85

Note.—For all F ratios, within-cells variance was used as the error term, with 8 *df*.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

effects of authoritarianism and its interactions with reward and prestige on number of communications initiated by Negroes. There were 10 main effects with probabilities below .10, including 7 without interactions (of which 4 were significant at $p < .05$). In general, the main effects depict Negroes in authoritarian groups as both more assertive and more cooperative than Negroes in nonauthoritarian groups. It appears that Negroes were more strongly affected in overt behavior by the F Scale scores of white *S*s than were whites themselves. Hence, we must assume that important behavioral cues were presented to Negro men by their white companions which our observational categories did not pick up.

Main Effects of Authoritarianism on Social Influence

Table 3 presents values of F for main effects of authoritarianism and its interactions with reward and prestige on the amount of change in *S*s' opinions under group influence. For the two group mazes each *S*'s change score was the total number of times his private prediscussion opinions about moves differed from subsequent unanimous group decisions. His score on the two human relations tasks was the sum of differences between his private prediscussion position on an attitude scale and the position later agreed on unanimously by the group. Table 3 shows that on both types of task there were trends ($p < .10$) in which

TABLE 2

VALUES OF F FOR MAIN EFFECTS OF AUTHORITARIANISM AND INTERACTIONS WITH PRESTIGE AND REWARD ON NUMBER OF COMMUNICATIONS INITIATED BY NEGROES

Task and Category	Main Effects		Interaction		
	Author. Higher	Non-author. Higher	A/P	A/R	A/P/R
Construction					
Gives help	5.33**		.12	.93	2.90
Asks for help	.47		.92	16.72***	.63
Orients group	.14		.61	.50	1.08
Human Relations					
Asks for info., etc.	5.90**		.19	1.54	.55
Disagrees	5.71**		.71	.67	.14
Gives info., etc.	5.25*		.23	.88	.27
Agrees	.66		1.46	1.01	.78
Hanfmann-Kasanin					
Gives opinion	14.07***		.98	15.05***	2.02
Disagrees	2.25		.16	8.72**	.38
Orients group	1.85		1.65	2.10	.62
Asks for opinion		.79	.47	2.70	.96
Spiral—Mixed Pairs					
Proposes action	21.63***		1.75	6.93**	3.87
Praises other	20.20***		.38	6.73**	4.83*
Accepts proposal	1.65		3.32	.64	.18
Expresses hostility	.40		.93	.21	.35
Rejects proposal		4.92*	.81	.40	.25
Spiral—Negro Pairs					
Proposes action	10.91**		.33	3.26	2.31
Accepts proposal	3.83*		1.82	1.82	.69
Praises other	.88		2.24	1.84	.20
Expresses hostility		1.32	.47	.19	.53
Rejects proposal		2.00	2.61	2.61	.96

* $p < .10$.

** $p < .05$.

*** $p < .01$.

TABLE 3

VALUES OF *F* FOR MAIN EFFECTS OF AUTHORITARIANISM AND INTERACTIONS WITH PRESTIGE AND REWARD ON AMOUNT OF OPINION CHANGE DUE TO GROUP INFLUENCE

	Main Effects		Interaction		
	Author. Higher	Non-author. Higher	A/P	A/R	A/P/R
Human Relations					
Negroes	.22		.88	.53	.26
Whites	4.70*		.44	.79	1.01
Mazes					
Negroes	.09		.30	1.02	.77
Whites	4.38*		.19	.56	1.82

* $p < .10$.

authoritarian white *S*s shifted more than nonauthoritarian whites, while Negroes seemed to be unaffected by the experimental conditions.

Main Effects of Authoritarianism on Personality Ratings

Group members rated one another on 11 trait variables at the termination of their group experience. The complete list of trait variables appears in Table 4. Scores on each variable, as well as two combinations of scores, were submitted to analysis of variance tests. Table 4 shows values of *F* for main effects of authoritarianism and its interactions with reward and prestige on ratings by white *S*s. There were 2 significant main effects ($p < .05$) of authoritarianism on white *S*s' ratings of one another, out of a total of 13 single and combined ratings. Both occurred without significant interactions. Authoritarian whites rated one another more "modest" and less "frank" than did nonauthoritarian whites. On white ratings of Negroes there were four significant main effects and one trend. Authoritarian whites rated Negroes higher on "bright," "mature," and "dominating" than did nonauthoritarian whites; they also gave Negroes a higher Total Ascendancy score, and tended to rate them higher on "leads." These results are consistent with findings presented above on the communicative behavior of Negro *S*s.

Negro ratings of group members were not influenced by the experimental variables, except for two trends toward main effects of authoritarianism: Negroes rated nonauthoritarian whites as brighter ($F = 4.00$; $p < .10$) and less cooperative ($F = 3.55$; $p < .10$).

Results on Productivity and Other Dependent Variables

Productivity was measured for the group mazes (number of errors), construction tasks

TABLE 4

VALUES OF *F* FOR MAIN EFFECTS OF AUTHORITARIANISM AND INTERACTIONS WITH PRESTIGE AND REWARD ON WHITE SUBJECTS' RATINGS OF OTHERS' TRAITS

Item	Ratings of Whites				Ratings of Negroes			
	Author. Higher	Non-author. Higher	A/P	A/R	Author. Higher	Non-author. Higher	A/P	A/R
Modest-conceited	14.47***			2.85		.48	.43	.80
Warm-cold		.95	6.60**	.42		.93	4.16*	.64
Bright-dull		.55	3.77*	.12	8.75**		.82	.48
Relaxed-tense	.31		.87	.23	3.26		.19	.16
Good mixer-poor mixer		.81	12.23***	3.31	1.85		.59	.36
Mature-immature	1.11		5.08*	.50	5.44**		4.00*	9.00**
Frank-concealing		8.25**	2.50	.38		.29	.91	.20
Cooperative-uncooperative	1.27		6.79**	4.45*	1.35		.35	.13
All Evaluative scales (above)		.69	13.89***	.30	1.59		2.18	.33
Dominating-submissive		1.43	.99	.69	7.33**		.84	7.33**
Leads-follows		.60	5.11*	.51	4.33*		.62	.25
Active-passive		1.13	8.00**	.87	2.82		.93	.61
All Ascendancy scales (above)		1.80	3.54*	.66	7.16**		.45	.45

Note.—Triple interactions are omitted. None were significant.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

(accuracy of models), the Hanfmann-Kasanin problem (solution time), and the ball-and-spiral (number of times various levels were reached, weighted for relative difficulty of levels). On the spiral task pairs of nonauthoritarian whites achieved significantly higher scores than did pairs of authoritarian whites ($F = 9.76$; $p < .05$). To save space other F ratios for productivity are not presented, since there were none that corresponded to $p < .10$.

Similarly, findings on the Defensiveness scale, the Satisfaction scale, social preference, and evaluations of performance are omitted because there were no values of F at $p < .10$.

Discussion of Main Effects

Authoritarian whites tended to behave more compliantly in biracial situations and evaluated Negroes more favorably than did nonauthoritarian whites, while Negroes were more ascendant with authoritarians than with nonauthoritarians. Our results are understandable in terms of the low perceived power of authoritarians in the present situation. The unbiased behavior of the white *E*, the university setting of the experiment, and the numerical and intellectual equality of white and Negro *S*s may have introduced strong restraints against open expression of anti-Negro sentiments. The fact that authoritarians were actually more deferential with Negroes than were nonauthoritarians, suggests the presence of internal controls so powerful as to produce behavior opposite to their hostile impulses. In line with this interpretation, there was no evidence of greater cooperativeness or more favorable trait ratings *between* authoritarians than *between* nonauthoritarians. Indeed, authoritarians' ratings of one another as more "modest" and less "frank" suggest awareness of a cautious, tactical quality of behavior. And when paired together on the spiral, authoritarians were less able to cooperate than were nonauthoritarians, as shown in lower achievement scores.

Interactions of Experimental Variables on Communication and Trait Ratings

White scores. In Table 1 it can be seen that there were six significant interactions of authoritarianism and prestige on white communication frequencies. And in Table 4 there are five significant interactions and six trends to interaction of these two variables on white *S*s' descriptions of other group members. Between-cell comparisons by *t* tests were not made because of the smallness of cell samples. However, inspection of cell means revealed perfectly consistent patterns in which scores on cooperative behavior and favorableness of personality ratings were always relatively high in the authoritarian-prestige treatment and in the nonauthoritarian-no prestige treatment, and were always relatively low in the other two treatments.

That authoritarian whites tended to respond more favorably to prestige than to no prestige is consistent with the general theory of the authoritarian personality. But it is unclear why nonauthoritarian *S*s tended to respond better to no prestige than to prestige. Possibly the ego-defensiveness of men in high prestige groups that was noted by Back (1951) and Katz et al. (1958) is especially strong in low F Scale scorers, who presumably are self-critical and would not readily accept an arbitrarily high appraisal of their ability (such as was given in this experiment by *E*).

A total of three significant triple interactions (authoritarianism × prestige × reward) on white communication scores had no consistent patterns of cell means.

Negro scores. There were five significant interactions of authoritarianism and reward on Negro communication frequencies (see Table 2). Inspection of cell means revealed that Negroes in groups with authoritarian whites were always more active under individual reward than under group reward, while Negroes in nonauthoritarian groups showed no consistent differences in behavior or ratings under the two reward conditions. We conjecture that Negroes perhaps were able to sense the presence of inhibited hostility in authoritarian whites, hence, tended to withdraw further from them in the condition of relatively high interdependence (group reward) than in the condition of relatively low interdependence (individual reward); but sensing little or no hostility in nonauthoritarian whites the Negro *S*s behaved about the same under both conditions of reward.

TABLE 5

VALUES OF *T* FOR RACIAL DIFFERENCES IN AMOUNT OF COMMUNICATION

Category	Task				
	Construction	Maze	Human Relations	Blocks	Spiral[a]
Gives help	1.79				
Asks for help	3.52***				
Gives info., etc.		3.63***	.65	1.58	
Asks for info., etc.		1.87*	1.46	1.50	
Agrees		.49	3.56***		
Disagrees		4.56***	4.45***	2.32**	
Proposes action					.10
Accepts proposal					1.95*
Rejects proposal					.23
Orients group	2.70**			.91	
Praises other					1.99*
Expresses hostility					1.58

Note.—For all values of *t* above 1.0 white scores were higher than Negro scores.

[a] Negro-white pairs only.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

Racial Differences for all Conditions Combined

Katz et al. (1958) found that for all experimental conditions combined, white *S*s, who were unselected with regard to F Scale scores, proved to be more active, communicative, and productive than their Negro companions, and tended to talk mainly to one another, while Negro *S*s oriented compliantly toward the whites. In the present study we wanted to see whether racial differences similar to those obtained previously would occur—this time on a wider range of dependent measures—when Negro and white *S*s were closely matched on intelligence and displayed equal mental competence (in terms of contrived scores).

To compare Negroes and whites on a given dependent variable a difference score was computed for each experimental group by subtracting the combined scores of the two Negroes from the combined scores of the two whites. For each distribution of difference scores a *t* test was used to evaluate the divergence of the mean from a theoretical value of zero. Since the 16 experimental groups had been subjected to eight experimental treatments (two in each treatment) the estimate of variance had 8 *df*.

Amount of communication. Table 5 shows, for each of 20 observational measures, the *t* value of the difference in number of communications initiated by Negroes and by whites. (Maze data are here included, since the sampling irregularities mentioned in the previous section did not invalidate Negro-white different scores.) Three trends and seven significant findings occurred, in all of which white scores were higher than Negro scores. Thus, on half of all the communication measures Negroes appeared to be less active than whites.

Direction of communication. S scores, as described by Katz et al. (1958), were used to measure the proportion of communications of various types that were directed to own-race persons and to other-race persons. These scores were not applied to spiral behavior, since in that situation only two men were present at a time. Two additional categories could not be scored for direction because of zero frequencies in some groups. In Table 6 are presented values of *t* for directional "bias"—i.e., deviations from random targeting of remarks—on the part of white and Negro initiators. With respect to white initiators, out of 13 scoreable categories, there were 5 on which the probability that targeting was random was less than .10. On 3 categories white *S*s favored one another as recipients, while on 2 categories they favored Negroes as recipients.

Table 7 shows that Negroes, on the other hand, spoke mainly to whites on 11 out of 13 categories (two trends and nine significant differences) and favored one another on none.

Social influence.[7] On the group mazes Negroes shifted more than whites from their private choice-point decisions to subsequent group decisions ($t = 2.44$; $p < .05$). The greater compliance of Negroes was not due to differences in maze learning ability, for there were no reliable differences in number of errors made by Negroes and by whites in their private decisions.

Social preference. On one of three sociometric items—preference as future working companions—Negroes favored one another ($t = 2.5$; $p < .05$), while whites showed no bias.

Rankings on task competence. At the end of the contact experience *S*s were required to

---

[7] For measures of social influence and all remaining dependent variables only results at $p < .10$ will be presented, in order to save space.

TABLE 6

VALUES OF t FOR TESTS OF "BIAS" IN DIRECTION OF WHITE COMMUNICATIONS

Category	Task and Favored Recipients							
	Construction		Maze		Human Relations		Blocks	
	White	Negro	White	Negro	White	Negro	White	Negro
Gives help	3.19**							
Asks for help	2.41**							
Gives info., etc.			.25		.22		.43	
Asks for info., etc.				1.09		3.83***	3.56***	
Agrees				1.74		2.07*		
Disagrees			1.28			1.67		1.05

* $p < .10$.
** $p < .05$.
*** $p < .01$.

rank one another for competence on the tasks they had performed. In analyzing these data we combined rankings on mazes, Hanfmann-Kasanin, and abstract reasoning into a single "cognitive tasks" score, while construction, spiral, and human relations were treated separately. Negroes ranked whites higher than they ranked each other on "cognitive tasks" ($t = 3.27$; $p < .05$) even though their objective performance was virtually the same as that of whites on the mazes and abstract reasoning and not reliably different on Hanfmann-Kasanin.

Satisfaction. Whites scored higher than Negroes on the Group Satisfaction scale ($t = 2.76$; $p < .05$).

Thus, even though Negro *S*s were closely matched with white group mates on intelligence and were given evidence of equal ability on cognitive tasks they maintained a compliant orientation toward the white men and evidenced relatively low satisfaction and self-confidence. White *S*s, on the other hand, gave no measurable indications of anti-Negro bias either in overt behavior or questionnaire responses.

TABLE 7

VALUES OF t FOR TESTS OF "BIAS" IN DIRECTION OF NEGRO COMMUNICATIONS

Category	Task[a]			
	Construction	Maze	Human Relations	Blocks
Gives help	.72			
Asks for help	.64			
Gives info., etc.		4.67***	5.04***	2.69**
Asks for info., etc.		2.29*	4.87***	2.27*
Agrees		4.43***	2.61**	
Disagrees		6.36***	3.09**	7.74***

[a] White recipients were always favored.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

SUMMARY

Sixteen groups consisting of two Negro and two white college students who were matched on intelligence worked under different combinations of group or individual reward and high or neutral group prestige. In half of the groups both white *S*s had high F Scale scores and in the other half both white *S*s had low F Scale scores. Measures of communication, social influence, productivity, and social perception were obtained.

Comparing authoritarian whites and non-authoritarian whites: on a task that required close cooperation between two *S*s authoritarians accepted significantly more suggestions from Negro *S*s and rejected significantly fewer; on two tasks requiring group decisions authoritarians showed trends toward greater compliance; authoritarians rated Negroes significantly higher on intelligence, maturity, and dominance; authoritarians were significantly less productive when paired together on a two-man task requiring cooperation. The first three effects were interpreted as due to the authoritarian's fear of revealing anti-Negro attitudes in a potentially punitive environment.

Negroes who worked with authoritarians

were generally more assertive and cooperative than Negroes who worked with nonauthoritarians, as indicated by trends or significant differences on 10 out of 21 communication categories.

For white *S*s there were trends or significant interactions of authoritarianism and prestige on 6 out of 21 types of communication frequencies and on 11 out of 26 ratings of others' traits. In all of these interactions inspection of cell means revealed that behavior was more cooperative or ratings were more favorable in the authoritarian-prestige and nonauthoritarian-no prestige treatments than in the nonauthoritarian-prestige and authoritarian-no prestige treatments.

For Negroes there were significant interactions of authoritarianism and reward on five types of communication frequencies, and inspection showed that Negroes grouped with authoritarian whites were always more active under individual reward than under group reward, while for Negroes in nonauthoritarian groups there were no consistent differences in behavior under the two conditions of reward.

For the entire sample of 16 groups it was found significantly that whites spoke more than Negroes, Negroes spoke more to whites than to one another, Negroes were more susceptible to group influence than were whites, ranked whites higher on mental ability, favored one another as future work companions, and scored lower than whites on a Group Satisfaction scale. These results indicate that even when Negroes are given objective evidence of equal mental ability in a relatively brief interracial contact they tend to feel inadequate and to orient compliantly toward whites.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSA, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BACK, K. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. *A manual for the Differential Aptitude Tests.* New York: Psychological Corp., 1947.

CHRISTIE, R., & COOK, PEGGY. A guide to published literature relating to the authoritarian personality through 1956. *J. Psychol.*, 1958, **45**, 171–199.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F Scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

FRENCH, J. R. P., JR. The disruption and cohesion of groups. In D. Cartwright & A. Zander (Eds.), *Group dynamics.* Evanston, Ill.: Row, Peterson, 1953.

GIBB, J. R. Factors producing defensive behavior within groups. Quarterly Status Report, Feb. 15, 1954, Contract Nonr-1147(03) NR 170-226.

KATZ, I., GOLDSTON, JUDITH, & BENJAMIN, L. Behavior and productivity in bi-racial work groups. *Hum. Relat.*, 1958, **11**, 123–141.

SMITH, E. E. Effects of threat induced by ambiguous role expectations on defensiveness and productivity in small groups. Technical Report No. 1, Aug. 1956, Contract Nonr-1147(03) NR 170–226.

THIBAUT, J., & RIECKEN, H. W. Authoritarianism, status, and the communication of aggression. *Hum. Relat.*, 1955, **8**, 95–120.

WEBSTER, H., SANFORD, N., & FREEDMAN, M. A new instrument for studying authoritarianism in personality. *J. Psychol.*, 1955, **40**, 73–84.

(Received September 28, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 457-462

SIZE ESTIMATION OF PICTURES ASSOCIATED WITH SUCCESS AND FAILURE AS A FUNCTION OF MANIFEST ANXIETY[1]

THEODORE P. ZAHN
National Institute of Mental Health

An unsettled problem of considerable theoretical interest concerns the relationship of attitudinal and motivational variables to judgments of magnitude. In various studies perceptual accentuation of the sizes of objects has been found to be a function of positive monetary (Bruner & Goodman, 1947; Dukes & Bevan, 1952) and symbolic value (Bruner & Postman, 1948), negative monetary (Dukes & Bevan, 1952) and symbolic value (Bruner & Postman, 1948), and positive reinforcement (Lambert, Solomon, & Watson, 1949). Granting the soundness of these empirical findings, the generalization is tenable that the personal relevance or importance of the object or symbol judged was the important variable underlying all of the above experiments.

A method of increasing the personal relevance of an object might be to make that object central in a task on which the subject (*S*) fails. If negative reinforcements occur in close conjunction with responses to the object, the latter may acquire secondary reinforcing properties so that it will evoke the same kind of emotional reaction as is produced by the negative reinforcement itself. If the above considerations are true the size of the object so reinforced should be overestimated.

In a previous study (Zahn, 1959), it was found that certain schizophrenic patients (i.e., those with good premorbid histories) accentuated the sizes of a picture associated with failure and of a picture of a mother scolding a boy relative to more positively toned pictures. It was hypothesized there that the patients showing this effect were characterized by more affective sensitivity or anxiety than were the patients who did not show the effect and than the normal controls who gave equivocal results.

The present experiment investigates the size estimation, from immediate memory, of pictures that have been associated with success and failure in a prior task by *S*s with varying degrees of manifest anxiety. The size estimations of these two pictures are compared with each other and with estimations of the sizes of pictures which have *not* been involved in a prior task. It was hypothesized that the failure picture would be overestimated relative to the success and neutral pictures primarily by *S*s with a high degree of manifest anxiety.

METHOD

The general procedure consisted of three parts: a session of size estimation during which the sizes of three pictures were judged, four trials per picture, by the method of average error; a discrimination task in which success and failure experiences were provided for two additional pictures; and another session of size estimation in which the two "treated" pictures and two others were judged as before.

One of the purposes of the experiment was to record psychophysiological measures (GSR and heart rate) during the procedure. Hence, the experiment was introduced as a study to determine how such measures behaved while *S*s were performing simple tasks. Four electrodes were secured to *S*'s left hand and arm and right leg during the experiment. The data thereby secured are not reported here, however.

Subjects. *S*s were 48 undergraduate students at Duke University. Of these, 29 "volunteered" for credit from sections of the spring semester course in elementary psychology and 13 from the summer session course. Six additional *S*s were paid for their participation. Ten of the *S*s were women.

The Taylor Manifest Anxiety Scale (MAS), plus 15 items from the MMPI Lie (*L*) scale, had been given to the psychology classes from which 42 *S*s were drawn. The six *S*s not taking psychology were administered the scale just before the experiment. The experimenter (*E*) was unaware of the scores of the *S*s at the time of the experiment. *S*s from the entire range of MAS scores were used. For most of the analyses in which two MAS groups were considered, the low anxious (LA) *S*s had scores of 4–13 and the high anxious (HA) *S*s 14–32. The results for three additional *S*s were discarded because of *L* scores of six or more.

The size estimation task. The apparatus for size estimation has been described in detail elsewhere (Harris, 1957). It allowed for the projection of an outline drawing (white lines on dark background) onto a translucent screen of tightly stretched draftsman's parchment large enough to minimize framework cues.

---

[1] This study was carried out at Duke University and supported by USPHS NIMH Grant M-629. The author wishes to express his sincere thanks to Eliot H. Rodnick for his suggestions and encouragement.

The size of the image on the screen could be adjusted by means of a knob protruding below the screen. The image remained in focus and constant in brightness at all sizes.

S estimated the sizes of the pictures by the method of adjustment. On each trial, the picture was exposed at the standard size for about 5 seconds by count; then it was shut off for about 4 seconds and re-exposed at a different size. *S* then adjusted the image to what he estimated to be the standard size. Four such trials were run on each picture.

The stimulus pictures were as follows: (*a*) a square; (*b*) two business buildings, one taller than the other (Bl); (*c*) a tree and a bush (Tr); (*d*) a house and a tree (HT); (*e*) a lamp and a table (LT); (*f*) a "feeding" picture of a "mother" about to pour a glass of milk out of a pitcher for her "son" who is holding a glass in his outstretched hand (Fe); and (*g*) a "scolding" picture of a "mother" with her arm raised in a scolding manner over a boy who stands with head down (Sc). *S*s were asked to imagine the most unpleasant instance involving being scolded by their mothers that they could while judging this picture.

All pictures were simple line drawings. Just before *S* was shown the picture for either of the two tasks, *E* told him which picture he would see with essentially the same description as presented here. Bl and Tr are referred to as "control" pictures. HT and LT were the pictures associated with success and failure as described below.

The height of the tallest object in each picture at the standard size was 16.5 centimeters on each trial. The sizes (height) at which the pictures were re-exposed were 13.5, 18.9, 21.2, and 14.8 centimeters in that order, constant for all *S*s and for all pictures judged.

The discrimination task. The discrimination task was used solely as a means for creating a success and a failure experience in association with different previously neutral pictures. The instructions stated that *E* was interested in how well *S* could tell the difference between two pictures when they were presented rapidly and that the crucial difference would be in some detail in the picture, *not* size or brightness.

The pictures were exposed by a Revere automatic projector, the gears of which were modified so as to allow for an exposure of approximately .20 second. On each trial the standard picture was exposed for about 2 seconds, then 2 seconds later the variable picture was flashed for .20 second. *S* indicated "same" or "different" by operating a Mallory switch which lit up either of two lights for *E*. No time requirements were put on this response. About 5 seconds after *S*'s response, *E* reinforced it by lighting up one of two boxes facing *S* just below the screen. The word "right" was illuminated in the aperture of one box in white letters; the word "wrong" in red letters was illuminated in the other box.

There were 30 such trials, 15 on each of two pictures. On the success picture, the standard and variable exposures differed in a small but discriminable way on 10 of the 15 trials; and *S*, regardless of his response, was flashed "right" on 13 of the 15 trials and "wrong" on 2 trials. On the failure picture, however, the standard and variable exposures were always *identical*; and *S* was flashed "wrong" on 9 or 13 of the trials, depending on the condition (see below), and "right" on the remainder.

The two pictures so reinforced were the HT and LT pictures. On variations of the HT picture, a limb of the tree varied in angle, and on the LT picture, a leg of the table varied in angle. These pictures were exposed at exactly the same size as the standard size of the size estimation procedure.

There were three conditions (for three groups of *S*s), varying in the sequence of pictures or the number of "wrongs" and "rights." In Condition 1, trials on the two pictures were mixed in a fixed irregular order, and the last 10 trials of the failure picture were uniformly reinforced by "wrong." Condition 2 was identical to Condition 1 except that on the last 10 trials of the failure picture, 7 of *S*'s responses were reinforced by "wrong" and 3 by "right." In Condition 3, trials on the success and failure pictures were alternated in a regular ABAB order, and the reinforcement schedule was the same as in Condition 2. This procedure was undertaken mainly because of the presumed effects on the psychophysiological measures.

For each of these conditions, approximately half the *S*s were successful on the HT picture and judged the size of it first; and half were successful on the LT picture and judged it second. These will be designated as orders S-First and F-First, respectively. Although this procedure confounds the effects of the order of judgment with the picture succeeded or failed on, it serves to counterbalance the effects of both.

Procedure. In the first session of size estimations, *S* judged the sizes of the square, either the Tr or Bl pictures (counterbalanced), and the Fe picture in that order. After the discrimination task, the two reinforced pictures (HT and LT) were judged, followed by whichever of Tr or Bl that was not previously judged. In order S-First, the Bl picture was judged in the first session and the Tr picture in the second, whereas in order F-First, the sequence of neutral pictures was the reverse.

Treatment of data. The data are expressed as the deviation of the adjusted size from the standard size in terms of the distance of the slide away from the light source in mm. This measure was used in an attempt to approximate the discriminability function for size. It very closely approximates a logarithmic function of the linear dimensions of the image on the screen and, hence, should be closely related to discriminable size differences. The score for any *S* on any one picture is the mean of four of these deviations.

RESULTS

Size Estimation and Total MAS Scores

Table 1 shows means and variances of the size estimation scores for the two anxiety groups on the two reinforced and nonreinforced pictures. All mean scores represent *underestimations* of the size of the picture compared to the standard size, the more negative the score, the greater the underestimation.

A trend for generally more underestimation in the HA group is indicated except in the case of the failure picture which is relatively overestimated by this group. This picture departs from the LA group trend in the direction of greater underestimation.

The most sensitive and least contaminated test of the hypothesis that success and failure experiences influenced the size estimations of the two anxiety groups differentially is to compare differences in this respect between the two pictures for each *S*. A preliminary analysis of these difference scores, which included the three conditions under which the discrimination task was administered, two orders of size estimation, and the two levels of anxiety as independent variables, showed significant ($p < .01$) effects attributable to the anxiety variable, but Bartlett's test revealed significant heterogeneity of variance. Since no transformation could be found to ameliorate this situation, the discrimination task conditions variable was collapsed. This source of variation entered into no significant relationships with the dependent variable and is perhaps of scant theoretical interest. Bartlett's test on the resulting four variances indicated no significant heterogeneity of variance ($p > .10$). The 2×2 analysis of variance carried out on the collapsed data[2] yielded significant effects associated with anxiety ($F = 8.04$, $df = 1, 43$, $p < .01$) but nonsignificant effects due to order and to the interaction ($p > .20$). Individual *t* tests of the overall means of the two groups indicate that the HA *S*s show a significant overestimation of the failure picture relative to the success picture ($t = 2.84$, $df = 22$, $p < .01$), but the relative underestimation trend for the LA *S*s is not significant ($t = 1.28$, $df = 24$, $p > .05$).

Since the *S*s in this experiment had MAS scores from essentially the whole range, it was possible to ascertain the magnitude of the relationship between MAS and success-failure scores for size estimation. For the S-First condition, the Pearson *r* between these two variables was $-.47$ ($p < .05$); for the F-First order, it was $-.25$ (NS). The pooled correlation (obtained by averaging Fisher's *z* scores) was $-.38$.[3] This significant ($p < .01$) but low correlation is slightly higher than those reported between MAS scores and eyelid conditioning (Hilgard, Jones, & Kaplan, 1951; Spence & Taylor, 1953). Visual inspection of the scatter plots revealed a slight curvalinearity, so correlation ratios (eta) were computed. These turned out to be .67 and .54 for the two conditions. A test for linearity of regression (Walker & Lev, 1953, p. 278) revealed no significant curvalinearity, indicating that the *r*'s reported above are probably fairly good estimates of the true relationship.

TABLE 1

SIZE ESTIMATION MEANS OF THE REINFORCED AND NONREINFORCED PICTURES BY HIGH AND LOW ANXIOUS SUBJECTS

Anxiety Group	Reinforced Pictures		Nonreinforced Pictures	
	Success	Failure	Building	Trees
High	−.181	−.082	−.173	−.155
Low	−.099	−.149	−.081	−.104

In order to test for overall differences in size estimation between the two anxiety groups and for differences between the reinforced and control pictures, the data in Table 1 were subjected to an analysis of variance. The only significant source of variation was the interaction involving pictures and anxiety groups ($F = 2.89$, $df = 3,128$, $p < .05$). Tests of the significance of the differences between the means of individual pictures for each MAS group reveal that there is a reliable difference only between the success and failure pictures for the HA *S*s ($t = 2.18$, $df = 24$, $p < .05$), although the difference between Failure and Bl pictures for this group approaches significance ($p < .10$). A separate analysis of variance for the control pictures alone shows no significant differences between groups.

The results for the Fe and Sc pictures were analyzed separately by *t* tests because the order of presentation of these pictures was not counterbalanced. These show that the size of

---

[2] Two cells had $N = 12$, one $N = 11$, and one $N = 13$. The mean of the cell was added to the smallest cell and the score closest to the mean of the largest cell was subtracted from that cell. One *df* was subtracted from total and from the error term in the analysis of variance.

[3] The correlation between these two variables computed for both conditions pooled was $-.31$. This likely underestimates the population value because of the difference between the two conditions in mean success-failure scores.

TABLE 2

SUCCESS MINUS FAILURE SIZE ESTIMATION SCORES AS A FUNCTION OF PSYCHOLOGICAL AND SOMATIC ANXIETY

Somatic Items	Psychological Items			
	High		Low	
	N	M	N	M
High	11	−.032[a]	8	.098
Low	15	−.162	14	.071

[a] A negative score indicates the failure picture was overestimated relative to the success picture; a positive score indicates the reverse.

the Sc picture was not judged differently from any of the others for either group of *S*s, nor was the group difference significant for this picture. The only significant difference involving the Fe picture was its relative overestimation in comparison to the success picture for the HA *S*s ($p < .01$). In view of the multiple comparisons involved and the lack of counterbalancing of order, this difference must be considered only suggestive.

Somatic and Psychological MAS Items

Examination of the MAS reveals at least two classes of items: those relating to physiological or somatic manifestations of anxiety and those relating more to purely "felt" or psychological anxiety. In an attempt to increase the precision of the instrument in predicting size estimation performance, the *S*s were scored separately on each of these two kinds of items. The 15 somatic items referred to gastrointestinal symptoms, headaches, blushing, sweating, etc.,[4] while the 35 psychological items referred to nervousness, fears, worries, self-esteem, etc. The *S*s were divided into two groups on each of these subscales. For the psychological items, the highs were those with a score of 10 or greater; for the somatic items, the highs were those with a score of 4 or greater. Table 2 shows the size estimation success-failure difference scores and Ns according to this twofold classification. A test for independence on the distribution of frequencies yielded a chi square of less than 1, suggesting a lack of relationship between the two sets of items. An analysis of variance was carried out on the size estimation scores, using a method for unequal cell frequencies (Walker & Lev, 1953). The F for the psychological items was 13.00, significant below the .001 level ($df = 1, 44$). The F for somatic items was 2.43, $.20 < p < .10$, but the difference is in the opposite direction from that expected, the somatic highs showing a relative overestimation of the success picture and the somatic lows showing the opposite trend. The interaction was not significant ($F = 1.05$). Correlations between the psychological anxiety score and success-failure difference were $-.62$ ($p < .005$) for the S-First order and $-.35$ ($p < .10$) for F-First, and the pooled correlation was $-.50$ ($p < .001$).

This analysis indicates that the discriminative power of the MAS with respect to the size estimation task of the present experiment lies in the items referring to psychological anxiety and not in the somatic items.

DISCUSSION

The results of this experiment reveal a difference between high and low manifest anxiety groups in their size estimations of pictures which had been associated with success and failure. This difference was due primarily to a relative overestimation of the failure picture by HA *S*s. Assuming that having failed in a task involving a certain stimulus may increase the "personal relevance" or "importance" of that stimulus for *S*, the data strengthen the generalization that such relevance is an important variable influencing the estimated size of the stimulus. The fact that only the HA *S*s evidenced a significant difference in estimated size between the two pictures indicates that the affective meaning of the failure picture may have been stronger for this group.

The data are also consistent with the suggestion made previously (Zahn, 1959) that a higher anxiety level may have been responsible for the relative overestimation of a failure picture by the good premorbid schizophrenic patients. It is interesting to note, however, in view of the similarities of method in the two experiments, that these patients show a much greater difference between the success and failure pictures than do the present HA normals. The difference on the same pictures for the patient group is −.272 for eight trials of size estimation and −.367 for the first four trials as compared to −.099 for the HA normals for four trials. This greater difference for the patients would seem to indicate a greater degree of anxiety or perceptual lability or both. In addition, the failure to find any overestimation of the Sc picture in the HA normals—despite their relative overestimation of the failure picture—emphasizes the overestimation of this picture by the good premorbid patients. This would seem to indicate either that normals are not sensitive to symbolic representations of traumatic events as compared to schizophrenic patients, or that the two groups in fact differ in their maternal relationships.

---

[4] These items are Nos. 4, 5, 11, 18, 24, 25, 27, 33, 36, 37, 38, 41, 43, 44, and 48 in Taylor (1953, p. 286). Revised forms of items were used (p. 288).

A subsidiary result of this study suggests there may be at least two independent classes of items in the Taylor Scale—here called "somatic" and "psychological" items—which may be measuring different aspects of anxiety. Although in the present case the separation of these items was rather arbitrary and based only on their face validity, the results suggest that the psychological items may carry most of the weight in predicting the effects of failure on performance. Certainly, they suggest a closer look at the individual items of the scale, possibly with factor or cluster analytic methods.

In view of other studies which have found a relationship between positive value and positive reinforcement and size estimation, the question of why the success picture was not also overestimated seems pertinent. One can only speculate in an ad hoc fashion that the operations performed here did not lead to an increase in the value or affective meaning of this picture. Since no special attempt was made to make the discrimination task an ego involving one, it is unlikely that the easy success achieved on the success picture led to any great feelings of accomplishment or of having overcome a challenge, and hence the picture remained quite neutral.

Insofar as the size estimation technique used in the present experiment is comparable to that of previous studies relating size estimation and value, it may be argued that the phenomenon demonstrated is essentially similar. However, a serious question may be raised about the comparability of the present procedure to that traditionally used in the coin studies and others. It can be argued, rightly, that with the relatively long delay between presentations of the standard and variable pictures (4 seconds) in the present procedure, what is being measured is memory and not perception. In fact, if only perception were altered by the variables used here, no differences should be manifest because the adjusted image should also be perceived as larger to the same extent that the standard was.

Granting this to be true, it is still possible to argue that results obtained by the present method are comparable to the previous ones. Certainly, they are comparable to those of Dukes and Bevan (1952) and of Gilchrist and Nesberg (1952), both of which involved a short interval between judgments of a standard and a variable stimulus. However, even in studies in which the standard is physically present while *S* is adjusting the variable stimulus, it would seem probable, perhaps even necessary, that *S* would regard the two stimuli with separate fixations and that the variable would be fixated after the standard. Thus, there is necessarily a finite time interval, albeit a very small one, between fixation of the standard and variable. Of course, more than one comparison is permissible in this procedure, but this should not change the situation appreciably because in any case the variable should be fixated last. Thus, there is a likelihood that the present procedure and those others involving a time interval between presentations of standard and variable stimuli differ from the "traditional" procedure, where both are simultaneously present, mainly in a quantitative fashion: in the time interval between fixating the two. The present procedure may offer less optimal judgmental conditions, but does not differ in any essential way or measure a different psychological function from the traditional procedure.

This analysis advances the suggestion that in all previous work on this problem, it is possible that changes in the memory rather than the perception of the valued or personally relevant objects were being assessed. Perhaps a procedure that would allow for a less equivocal interpretation would successively present the standard stimulus as a neutral figure (i.e., square or circle) and the variable the value figure. If the value or personal relevance were "built-in" in a manner similar to the one used in the present study, structural properties of the stimuli could be controlled by proper counterbalancing. This method would seem inappropriate for stimuli with "inherent" value, such as coins, hammers and sickles, and mothers, however, unless some other kind of structural factors were utilized to control for differences between standard and variable stimuli.

SUMMARY

An experiment was performed to test the hypothesis that the size of a picture associated with failure would be overestimated relative to that of a picture associated with success and control pictures primarily by *S*s with high scores on the Taylor Manifest Anxiety Scale. After practice on a size estimation task, *S*s were given a discrimination task involving two pictures, responses to one of which were generally followed by "right" and responses to the other by "wrong." The sizes of these pictures and others were then judged by the method of average error.

The results showed that, in accordance with the hypothesis, high MAS *S*s significantly overestimated the size of the failure picture relative to that of the success picture, but the low MAS group showed an insignificant trend in the opposite direction. In addition, it was shown that the "psychological" anxiety items of the Taylor scale related significantly to the success-failure difference in size estimation, but the "somatic" items did not.

The results are a further confirmation of the generalization that personal relevance is a determinant of size estimation and imply that associating a stimulus with failure makes it more personally relevant for high anxious than for low anxious *S*s.

REFERENCES

BRUNER, J. S., & GOODMAN, CECILE C. Value and need as organizing factors in perception *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.

BRUNER, J. S., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.

DUKES, W. F., & BEVAN, W., JR. Size estimation and monetary value: A correlation. *J. Psychol.*, 1952, **34**, 43–53.

GILCHRIST, J. C., & NESBERG, L. S. Need and perceptual change in need related objects. *J. exp. Psychol.*, 1952, **44**, 369–377.

HARRIS, J. G., JR. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *J. Pers.*, 1957, **25**, 651–671.

HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–100.

LAMBERT, W. W., SOLOMON, R. L., & WATSON, P. D. Reinforcement and extinction as factors in size estimation. *J. exp. Psychol.*, 1949, **39**, 637–641.

SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

ZAHN, T. P. Acquired and symbolic affective value as determinants of size estimation in schizophrenic and normal subjects. *J. abnorm. soc. Psychol.*, 1959, **58**, 39–47.

(Received October 1, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 463–468

THE EFFECT OF AROUSED MOTIVATION ON ATTITUDES[1]

HELEN PEAK
University of Michigan

EARLIER experiments have demonstrated that attitudes toward certain situations, policies, or actions are predictable from information about the consequences believed to follow from or to be connected with those situations. Thus attitudes toward removing racial segregation were predictable from a combination of information about the perceived consequences (or instrumental properties) of desegregation and judgments of the value of those consequences (Rosenberg, 1956). Furthermore, attitudes toward this issue were changed by altering the instrumental relations (Carlson, 1956; Peak, Morrison, Spivak, & Zinnes, 1956). But if attitudes are thus dependent on the existence of relations between the object of the attitude (viz., Negro segregation) and positively and negatively valued consequences, then attitudes should change if something alters evaluation of those consequences. Suppose, for example, that a student's negative attitude toward his roommate derives in part from a belief that his roommate interferes with studying and thus with making high grades. This attitude might be altered either by changing the belief that the roommate is responsible for the study difficulties (change in perceived instrumentality) or by changing the evaluation of high grades.

The present experiment is an attempt to determine whether attitudes can be changed by intensifying consequences. A number of academic situations were chosen as the attitude "objects" and their perceived relation to making good grades was determined. Then an effort was made to arouse the goal of making good grades by the device of testing immediately after a quiz had been given (arousal condition), on the assumption that the motive of achieving grades is activated more strongly on such an occasion than on an ordinary class day (neutral condition).

I have sketched elsewhere a behavior theory which assumes that the course of response on any occasion depends on a combination of the pattern of *input* at particular points in a psychological structure and on the characteristics of that *structure* (Peak, 1958). The theory also makes the familiar distinction between restructuring or learning and performance (Peak, 1959). Thus change in attitudinal responses may be produced by a modification in the system of relations (structure) on which the response depends, as when the relation of an attitude object to valued goals is altered (Carlson, 1956). Or activating conditions may be changed. The change in attitudinal response predicted in the present study is probably best thought of as the result of a change in input patterns or activation, i.e., in the degree of arousal of the need to make good grades. Whether any such activation results in restructuring can be determined only by testing reactions under activation conditions that were operating at the time that the original measures were made.

HYPOTHESES

The first hypothesis states that attitudes expressed toward any issue or object (X) are a function of the instrumental relations of X to valued consequences.

The second hypothesis is this. Expressed attitudes toward an issue (X) are a function of the intensity of arousal of the consequences of X, other things being equal. More specifically, attitudes toward any X, seen as positively instrumental to attaining good grades, tend to become more positive when there is an increased arousal of the desire to make good grades; with negative instrumentality, attitudes toward X should become less positive under arousal conditions, when positive and negative instrumentality groups are compared to each other and to those judging X to have no instrumental value.

---

[1] The analysis of these data was completed under contract between the Office of Naval Research and the University of Michigan, Project Nonr 1224(10) NR 171-039. The article was written while the author was in residence at the Center for Advanced Study in the Behavioral Sciences. We are grateful for the skillful assistance with computations by the late Mary Girschick of the center staff.

TABLE 1
ATTITUDES AND ATTITUDE CHANGE AS A FUNCTION OF INSTRUMENTALITY AND GOAL AROUSAL

Item	Sex	p (two-tailed) of Chance Relation						Instr. (Aroused) and Instr. (Neutral)		
		Instr. (Aroused) and Attitude (Aroused) (1)			Instr. (Aroused) and Attitude Change (2)			phi	p (3)	
1.	M	$<.01$	(59)[a]	Pre[b]	$>.10$	(59)	Pre	.35	$<.01$	(63)
	W	$>.10$	(101)	Pre	$>.10$	(100)	Pre	.49	$<.001$	(107)
	C				$<.02$	(159)	Pre			
2.	M	$<.001$	(65)	Pre	$>.10$	(65)	Cur[c]	.45	$<.001$	(63)
	W	$<.001$	(105)	Pre	$>.10$	(105)	Cur	.50	$<.001$	(108)
	C				$>.10$	(170)	Cur			
4.	M	$<.001$	(59)	Pre	$>.10$	(57)	Pre	.35	$<.10$	(63)
	W	$<.001$	(100)	Pre	$>.10$	(100)	Pre	.51	$<.001$	(105)
	C				$>.10$	(157)	Pre			
5.	M	$<.001$	(58)	Pre	$<.001$	(63)	Pre	.52	$<.001$	(64)
	W	$<.001$	(106)	Pre	$>.10$	(101)	Pre	.50	$<.001$	(109)
	C				$<.001$	(164)	Pre			
6.	M	$>.10$	(59)	Pre	$<.02$	(59)	Pre	.52	$<.001$	(64)
	W	$<.001$	(102)	Pre	$<.05$	(102)	Pre	.44	$<.001$	(108)
	C				$<.01$	(161)	Pre			
8.	M	$<.001$	(62)	Pre	$<.05$	(62)	Pre	.44	$<.001$	(61)
	W	$<.001$	(106)	Pre	$<.01$	(106)	Pre	.43	$<.001$	(104)
	C				$<.05$	(168)	Pre			
9.	M	$>.10$	(65)	Pre	$>.10$	(65)	Pre	.32	$<.02$	(64)
	W	$<.001$	(107)	Pre	$>.10$	(107)	Pre	.62	$<.001$	(109)
	C				$<.10$	(172)	Pre			
10.	M	$<.05$	(61)	Pre	$>.10$	(61)	Pre	.41	$<.01$	(61)
	W	$<.001$	(103)	Pre	$<.10$	(103)	Pre	.57	$<.001$	(107)
	C				$>.10$	(164)	Pre			
11.	M	$>.10$	(64)		$>.10$	(64)	Pre	.43	$<.001$	(62)
	W	$<.001$	(107)	Pre	$<.10$	(107)	Pre	.25	$<.02$	(105)
	C				$>.10$	(171)	Pre			
12.	M	$<.01$	(63)	Pre	$<.05$	(63)	Pre	.35	$<.01$	(62)
	W	$<.001$	(107)	Pre	$<.05$	(107)	Pre	.60	$<.001$	(107)
	C				$<.01$	(170)	Pre			
15.	M	$<.01$	(63)	Pre	$>.10$	(63)	Cur	.46	$<.001$	(62)
	W	$<.001$	(102)	Pre	$>.10$	(102)	Cur	.40	$<.001$	(104)
	C				$>.10$	(165)	Cur			

[a] Number of cases in parentheses.

[b] Pre indicates the relation was in the predicted direction.

[c] Curvilinear relation indicated by Cur.

tive and zero plus positive, or positive and zero plus negative). Results for 9 of the 11 items were in the predicted direction for men; the same number for women and the same number for men and women combined ($p <$ .033, one-tailed sign test). For men 4 of the 9 relations in the predicted direction were significant; none were significant which did not show predicted trends. For women 5 of the 9 were significant when one-tailed tests were used. When results of men and women were combined, 6 of the 9 items showed significant relations in the predicted direction using one-tailed tests. These findings suggest then, that goal-arousal conditions did tend to be associated with increased frequency of positive attitudes toward policies seen as positively instrumental to the goal and with decreased frequency of positive attitudes toward those issues negatively instrumental.

It may be concluded, therefore, that Hypothesis II was supported unambiguously for six of the items for the sample as a whole. Of the remaining five, three were in the predicted direction. The two which did not show predicted attitude change as a function of instrumentality judgments and of goal arousal were Items 2 and 15. The first of these stated that: "Giving unannounced quizzes is justified and should be general practice." The other was: "Studying at night is better than studying in the daytime." The reason for the deviant behavior of these particular items is not obvious.

Instrumentality under Aroused and Neutral Conditions

In the results reported, measurements of instrumentality made under arousal conditions have been employed. It is necessary, therefore, to report the relation between the two sets of instrumentality measurements.

Phi coefficients were determined from contingency tables for each item for men and for women and are presented in Column 3 of Table 1. It is clear that the correlations were all positive and significant. Probabilities are those of the contingency tables from which phi was derived.

Evidence of a possible shift in the direction of the instrumentality judgments was sought in the proportion of cases in which there was positive, negative, and no change in the instrumentality direction. Cases of no change were divided equally between the positive and negative categories and the departure of these frequencies from chance was determined.

In the case of Item 1 (taking books out of Library without permission) the instrumentality ratings became more negative under arousal conditions in men ($p < .10$). Ratings of Item 6 (no class should have more than 20 students) in women became more negative following arousal ($p < .05$); men changed in the same direction but not significantly. Women also saw Item 11 (8 hours of study) as less positively instrumental under arousal ($p < .05$). Men changed in the same direction but not significantly. Finally, Item 12 (maintaining good physical conditions) was judged less positively instrumental ($p < .05$ in men; $p > .10$ in women). The fact that all significant changes under arousal were in the direction of seeing the procedures as less positively effective toward the goal may mean that judgments made immediately after taking the quiz were more "realistic" and that some of the policies referred to in the items did not appear to be as helpful as had been anticipated. As a matter of fact, the correlations between attitude and instrumentality and between attitude change and instrumentality were somewhat higher when instrumentality judgments made under arousal conditions were used rather than those in the neutral condition.

F Scores

The three F Scale items and the items concerning the subjects' feelings of adequacy did not have any relation with the variables of this experiment.

SUMMARY

In each of two sessions 200 undergraduates judged the degree of agreement or disagreement with 11 academic policies. They then judged each of the policies (viz., having small discussion groups instead of lectures) in terms of the degree to which the policy in question either facilitated or interfered with the attainment of good grades; this was known as instrumentality sorting. At the beginning of each of these sessions, subjects reacted to Thematic Apperception Pictures which have been standardized as measures of need Achievement.

In one half of the total sample, the first session followed immediately after a 15-minute quiz which had been announced in advance. This was the arousal condition. For this group the second session 2 weeks or 10 days later was a neutral condition. In the other half of the sample, the neutral condition was established in the first session and the arousal condition in the second.

Hypothesis I stated that the attitude toward a policy or object is a function of the judged consequences of the policy. The hypothesis was strongly supported as it had been in earlier studies.

Hypothesis II stated that the attitude expressed is a function of the degree of arousal of consequences to which the attitude object has an instrumental relation. There were two approaches to the test of this hypothesis. It had been expected that scores on the need Achievement Test would reflect the effect of arousal conditions and that those who increased their achievement scores would show the predicted relation between instrumentality and attitude change more clearly than those whose achievement scores decreased or did not change as a result of arousal. The analyses produced no significant relations between change in achievement scores and experimental conditions or attitude change.

The other test of the hypothesis involved the observation of change in attitude toward a policy from neutral to arousal conditions as a function of judgments of the positive, zero, or negative instrumentality of the policy for making good grades. When tested by this method Hypothesis II was supported. The direction of attitude change was that predicted in a significant proportion of the items: i.e., when a policy was judged to be positively instrumental, attitudes toward it became more positive under arousal than when instrumentality was zero or negative. On the other hand attitudes tended to become more negative as a result of arousal when the instrumentality was negative.

This effect of arousal on expressed attitude is regarded as demonstrating the role of so-called performance or activation factors in determining attitude responses.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

CARLSON, E. R. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

PEAK, H. Psychological structure and psychological activity. *Psychol. Rev.*, 1958, **65**, 325–347.

PEAK, H. Attitude change and attitude structure. Final Report, 1959, University of Michigan, Contract Nonr 1224(10) NR 171–039.

PEAK, H., MORRISON, H. W., SPIVAK, M., & ZINNES, J. L. Factors in resistance to attitude change, personality factors. Technical Report No. 1, 1956, University of Michigan, Contract Nonr 1224(10) NR 171–039.

ROSENBERG, M. J. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, **53**, 367–372.

(Received October 2, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 469–473

AUDIENCE STRESS, PERSONALITY, AND SPEECH[1]

HARRY LEVIN, ALFRED L. BALDWIN, MARY GALLWEY,[2] AND ALLAN PAIVIO[3]
Cornell University

SPEECH appears to be an especially sensitive indicator of a person's feelings. Even excluding the content of what a person says, we often attribute a variety of traits or momentary states to him on the basis of such paralinguistic qualities as length of utterance, pauses, inflection, intonation, stress, etc. For example, the statement that a person is verbose connotes a host of personal characteristics, including that he welcomes a lengthy exposure to his listeners. Also, one of the most ubiquitous inferences from voice qualities is that the speaker is tense or nervous or ill at ease. Brown (1958) calls the bases of such inference the "expressive symptoms" of language.

The present research deals with two aspects of children's speech: the amount of time the child spends in talking and the number of speech errors he makes during his discourse. Our purpose is to predict each of these speech behaviors from situational factors—the number of people listening to the child talk—and from two personality dispositions which we have labeled "Exhibitionism" and "Self-Consciousness" (hereafter referred to as Ex and S-C).

Speaking in public may be thought of, generally, as a stressful situation. If people are forced into this condition and given the option of shortening or prolonging their public exposure, we may reasonably expect them to terminate their performance more rapidly than if the condition were less public. On the other hand, there are differences among people that influence how easily they tolerate public exposure and, in fact, to what extent they actually seek publicness. These individual variations complicate the simple situational prediction. The measures of Ex and S-C were designed to ascertain more precisely the ways in which the public exposure might be handled by various subjects.

The two dispositions may be thought of as approach and avoidance tendencies, respectively, to public performance. The Ex items deal quite specifically with public performance or the exposure of one's products to public appraisal. The briefer S-C scale taps a more general disposition of apprehensiveness about one's evaluation by others. The latter scale correlates substantially with such general anxiety scales as the Children's Manifest Anxiety Scale and Sarason's Test Anxiety Scale. The development of the two scales has been described in detail elsewhere (Paivio, Baldwin, & Berger, in press). They have been validated against volunteering for performance in skits, and their child rearing antecedents have been studied (Paivio, 1959).

If children are given the option of prolonging or terminating their public performance, highly exhibitionistic children for whom this occasion is pleasureable should extend their discourses compared to their less exhibitionistic peers. Conversely, self-conscious subjects, for whom talking to an audience is discomforting might be expected to exercise their option to endure such exposure as briefly as they can.

This prediction suggests that the two variables are negatively correlated. While they do have opposite effects on willingness to submit to public exposure, they are not the opposite ends of a bipolar scale.

Based on a number of samples of children, exhibitionism and self-consciousness have been found to be only minimally or not at all correlated. A child, thus, may wish to perform in public, yet be apprehensive about this exposure. This phenomenon is reminiscent of stage fright among actors. We reasoned, consequently, that subjects who score high on both scales would be experiencing an approach-avoidance conflict during their public performance—they want to get out of the situation and yet are drawn toward it. Such conflict may be contrasted with the less complicated apprehension felt by children who

---

[1] This study was supported by grants from the National Institute of Mental Health (M-901). We wish to thank Irene Stein, Katrin Norris, and Janet Von Hendy for their help with this study.

[2] Now at Washington State University.

[3] Now at the University of New Brunswick.

are very self-conscious but not exhibitionistic. A mode of resolution of the latter state is readily available to the subject: he can make his performance brief and thus leave the stressful situation. Such a straightforward termination of conflict is not available to the high-high group, because the act of leaving increases the positive valence of remaining before the public, which, in turn, calls out their self-consciousness, etc.

Besides influencing the length of their public performance, the conflict is also presumed to result in more speech errors. Since the subject who exhibits a high degree of approach and avoidance tendencies is under an unresolved stress situation in public, we expect his voice to exhibit more symptomatic signs of stress than the voices of subjects who have a different pattern of the two personality variables.

In summary, this study tests the hypothesis that on the average the public performance of speech is a stressful situation to children in our society and as such results in both briefer discourses and more expressive symptoms of uneasiness in the voice. The extent, and possibly the direction, of the influence of the public situation on a particular person, however, should depend on the degree to which he possesses two personality characteristics, Ex and S-C. Children high in Ex and low in S-C should find public speaking less anxiety provoking and more attractive than those with the opposite pattern. The subject who possesses a high degree of both of these dispositions should, however, be under greatest conflict during public performance, a state that should be reflected in a high rate of speech errors.

METHOD

Subjects

The subjects were 48 children, equal numbers of boys and girls, aged 10 to 12 years. They were all from a single elementary school in a lower middle class neighborhood.

Procedure

Measures of Exhibitionism and Self-Consciousness. Details about the nature of the items in the two scales and their factor analytic structure are given in detail in another publication (Paivio et al., in press). Ex consists of 13 items, each of which contributes a weight of +1 to the scale score. Examples of items that have high factor loadings on the scale are: "I like to tell a story in front of the class"; "I like to show things I make to other children."

The S-C scale has six items. Two examples of these items are: "I often wonder what others think of me"; "Other people hurt my feelings easily."

Test-retest reliabilities on these scales are .62 for Ex, and .45 for S-C, based on a sample of 209 children comparable in age and schooling to the subjects in the present study.

The items were embedded in a total of 79 and administered to 115 children in four classrooms. The final sample of 48 subjects was selected to represent four groups of boys and four of girls according to whether they scored above or below the medians for the total sample tested on the Ex and S-C scales. That is, one group was high Ex, high S-C, another high Ex, low S-C, etc.

The Public Speaking Situation. Each subject told stories under both "private" and "public" conditions. Before his first formal story the subject was trained in story telling in the following way. The experimenter took him from his classroom to another room where he was given a sentence and told to make up a story about it. During this preliminary session, the experimenter answered questions, asked the usual projective queries, and prompted the subject if he paused for so long that it seemed he would not go on without help. In order to help the child get the idea of a full story, the experimenter pointed out after the story was finished, that he had told a story which included a beginning, what happened, and how it came out.

Private and public experiences were counterbalanced so that following the preparatory session the child told two stories to the same experimenter or went to another room where he told two stories to an audience of six adults; the original experimenter was not present during this condition. The adults in the audience were instructed to listen interestedly and to respond naturally, but not to interact directly with the children. The two sessions for any single child were about one week apart. All stories were tape recorded.

Four beginnings used to instigate the stories were randomly assigned across children and conditions and were worded to match the sex of the child:[4]

1. A boy gets up in front of the class to make a speech.
2. A boy is a good actor. There is going to be a school play.
3. The teacher asks a boy to come to the front of the room and work at the blackboard.
4. A boy has prepared a long report on social studies. The teacher asks him to come to the front of the room and lead a discussion about the report with the children.

Two coders working independently, scored the following categories from the tapes:

1. Length of speaking (seconds)

---

[4] The story instigators, which are all obviously concerned with public performance, were developed in an earlier study to yield roughly comparable contents. The analysis of the content of the oral stories in the present study showed no consistent variation with either the personality or situational variables.

2. Number of words uttered
3. Number of stutters
4. Prolongations (like "sp-e-e-ech," or "my-y-y")
5. Shifts (starting a word or sentence and shifting before the utterance is finished, as "the b-girls want" or "the girls went—want")
6. Sighs
7. Pauses (any cessation of speech for 2 seconds or more was scored and the actual length of the pause was timed)
8. Vocal segregates (these are the ubiquitous "uhs" around words)
9. Catches in the voice

The lowest correlation between scorers on any one category is .80 and the highest is above .90.

In summary, this study involves a 2^5 factorial design made up of the following classifications: Ex, S-C, audience condition, sequence, and sex. The dependent variables are expressive speech characteristics symptomatic of stress, and length of voluntary exposure in story telling.

RESULTS

We are concerned with the prediction of two dependent variables—speech errors and length of speech—from personality and situational variables. The two dependent variables are uncorrelated in our data.

Length of Talk

Based on a previous study (Levin & Baldwin, 1958) and the pretests of this study we hypothesized that highly exhibitionistic subjects should speak for a longer period of time, given the opportunity, than their less exhibitionistic peers. Further, if we continue our assumption that the audience condition is more stressful than the more private one, and we recall that the child controls, to a great extent, the length of his exposure, we may predict briefer literary creations in the audience than in the private condition. This last anticipated finding may be qualified by interactions with the personality measures in that highly exhibitionistic subjects should be least discomfited by the public condition and highly self-conscious children, most.

The mean seconds speaking according to the various predictor categories are given in Table 1. From the analysis of variance the relevant findings are: Public ($F = 25.60$, $p < .001$); Public × Ex ($F = 4.78$, $p < .05$); Public × S-C ($F = 5.19$, $p < .05$).[5] Almost every one of the 48 children told briefer stories to the six listeners than to the single experimenter. Overlaid on this general decrease, however, is the finding that highly exhibitionistic children decreased their stories *least*. Subjects who are very self-conscious, on the other hand, showed the greatest shortening of their stories in the public compared to the single auditor condition. It will be recalled that the selection of the research groups obviates any correlation between the two personality measures. The two interactions with "public" summate so that the most spectacular decrease in time of exposure to the listeners occurred for the low Ex–high S-C group: a decrease of over 3 minutes, on the average.

TABLE 1
MEAN NUMBER OF SECONDS OF SPEAKING
(N = 12 in each cell)

Exhibitionism	Self-Consciousness	Condition: Public	Condition: Private
High	High	208	275
	Low	227	265
Low	High	141	341
	Low	168	231

Although we had no definite predictions about the highly self-conscious children who were not exhibitionistic, their long discourses in the private condition invite speculation. It may be that these usually reticent children overrespond to the warmth and individual attention of the private condition as it was designed in this study.

Speech Errors

For the two major personality variables—Ex and S-C—we predicted that the subjects who score high on both scales would be under greatest stress in a public situation and that the stress would produce a high incidence of speech mistakes. Further, the subjects who are highly Ex but not S-C should be least upset by public performance and therefore make few speech errors. Also, we anticipated a situational main effect indicating greater stress when the child was talking in public than when making up stories to the experimenter alone.

The means relevant to testing these hypotheses are presented in Table 2.

---

[5] Several other significant interactions that were not predicted but do not influence the interpretation of the above two findings are not being considered.

TABLE 2
MEAN RATES OF SPEECH ERRORS
(N = 12 in each cell)

Exhibitionism	Self-Consciousness	Condition	
		Public	Private
High	High	.071	.055
	Low	.032	.041
Low	High	.045	.057
	Low	.063	.056

The frequency of any one of the speech categories was usually so low that we calculated an overall speech error score by summing all of the categories and dividing the sum by the total number of words uttered, so that the score we are dealing with is "rate of speech errors." The distribution of this score was normalized by an arc sine $\sqrt{X}$ transformation (the means in Table 2 are not transformed).

An analysis of variance of these data indicated the following results: Ex × S-C ($F = 5.12$, $p < .05$); Ex × S-C × Public ($F = 5.14$, $p < .05$). The basic hypothesis is confirmed. Most speech errors are made by children who are conflicted over public performance, i.e., high in both Ex and S-C, and fewest errors are made by exhibitionistic children who show little apprehension about exposure (the high Ex and low S-C group). In addition, whether the child is performing before many or one observer puts a sort of astigmatic glass over this finding. In private, the strong implication is that only the high-low group is different from the others—they err least. In public, the conflicted (high-high) group is far higher in mistakes than the others.

When we look farther than the predicted error rate in the high-high group, the array of means in Table 2 raises some perplexing questions. The second largest error rate was made by subjects who are low on both characteristics. Our impression is that these children are generally unresponsive and apathetic. For them, speech errors—perhaps especially pauses—are not so much a response to situational stress as a generally slow behavioral pace. The rather close correspondence between the speech errors in both the public and private conditions favors our speculations. It suggests also, that future studies of paralinguistic behavior include a "normal" baseline derived from various natural speech situations.

In general, our hypotheses about speech errors and length of exposure are confirmed, with the exception that Ex did not appear as a significant main effect with seconds talking. It is interesting to note that when "number of words" rather than seconds is the dependent variable, highly exhibitionistic children utter more words irrespective of the audience condition.

DISCUSSION

Our theoretical assumptions at the outset of the study were that we might isolate two tendencies toward public performance. We thought of "exhibitionism" as an approach tendency toward such exposed situations and "self-consciousness" as an avoidance tendency. The tendencies are, in our theory, dispositions which are elicited when the person contemplates or is actually in a public situation.

These assumptions tend to be confirmed by the speech errors which we take as symptoms of stress. People who possess both these tendencies to a strong degree should be most conflicted (stressful) under an audience situation and this is actually the case. Also, the exacerbation of the larger audience magnifies the effect.

However, the results for length of exposure raise the question of whether there is ever an unadulterated seeking and preference for public performance. There is, for example, much folklore about experienced performers who repeatedly suffer agonies of stage fright. Similarly, even the highly exhibitionistic subjects told briefer stories in the public as compared to the private situation, although it is true that their reduction in time was considerably less than other children. The point, nevertheless, is that they did suffer less exposure when reciting to a group. Perhaps, the Ex variable might better be thought of as relative weights of approach and apprehension over public behavior.

Ex and S-C are conceptually related to a person's self-esteem. A person may score high on Ex because he sees the probability of success as high and just the converse for the self-conscious person. Yet for the exhibitionistic person a special kind of success may draw him to audience situations and that is success or achievement *in public.* This reasoning has lead us to make a distinction in conventional achievement motivation theory.

Many scorable instances of achievement according to McClelland, Atkinson, Clark, and Lowell (1953) involve achievement in public. In fact public approval for goal attainment appears to be a dominant motive for some people. It seems to us useful therefore, to make a distinction between pure achievement, where public performance is not relevant, and exhibitionistic achievement, where reaching the goal is simply instrumental to public approbation. We speculate that highly exhibitionistic children are such instrumental achievers. One clue to this interpretation is that they deliver longer stories, in number of words, in both audience conditions than do subjects who score at the lower end of this scale. The present findings set us to try to explicate the distinction between pure and exhibitionistic achievers in our further research.

SUMMARY

This study was designed to test the effects of a situational factor—the presence or absence of an audience—and two personality variables—exhibitionism and self-consciousness—on the length of speech and on the number of errors in speaking. Forty-eight children, aged 10 to 12 years, told two stories to the experimenter alone and two to an audience of six adults. The story beginnings were given to the subjects. The findings were:

1. Almost every subject told a longer story in the private compared to the audience situation.
2. Highly exhibitionistic children were least influenced in this respect whereas self-conscious children reduced the length of their stories drastically before the six observers.
3. Children who were both highly exhibitionistic and self-conscious made the most speech errors.
4. This finding was magnified under the audience condition.

REFERENCES

BROWN, R. *Words and things.* Glencoe, Ill.: Free Press, 1958.

LEVIN, H., & BALDWIN, A. L. The choice to exhibit. *Child Develpm.*, 1958, **29**, 373–380.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

PAIVIO, A. Child rearing antecedents of audience sensitivity. Unpublished doctoral thesis, McGill University, 1959.

PAIVIO, A., BALDWIN, A. L., & BERGER, S. Measurement of children's sensitivity to audiences. *Child Develpm.*, in press.

(Received October 7, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 474–479

CASE REPORTS

THE CONTEXTUAL ASSOCIATES OF SUN AND GOD IN SCHREBER'S AUTOBIOGRAPHY[1]

JULIUS LAFFAL
Veterans Administration Hospital, West Haven, Connecticut

DANIEL PAUL SCHREBER was a doctor of jurisprudence and a prominent judge at the end of the last century, who, after becoming psychotic, published an account (1903) of his delusions in the belief that he was making a contribution of value to science and the knowledge of religious truths. Without having known Schreber personally, Freud (1948) undertook on the basis of the published autobiography, to trace the etiology of Schreber's delusions, and in his study formulated the now famous paradigm of homosexuality and projection in paranoia. Interest in the Schreber case has continued high among psychoanalysts (Baumeyer, 1956) and a number of papers have been published on various aspects of the case. One notable study is that of Macalpine and Hunter (1955) who also published an English translation of Schreber's autobiography (1955). These authors differed with Freud's interpretation of the importance of homosexuality in the paranoid system, and maintained that a wish for ambisexuality and self-contained procreative powers was at the root of Schreber's illness. They brought the difference between their view and Freud's to a focus in contesting Freud's account of the symbolism of the sun in Schreber's delusions (Macalpine & Hunter, 1955):

> . . . if the sun was not a father symbol for Schreber, Freud's theoretical deductions only partially cover the facts. The passive homosexual wish-fantasy, ultimately derived from his infantile relation to his father and evidenced in his relation to the sun, could not afford the whole explanation, but would throw light only on one aspect, however important, of his psychosis and the content of his delusions.
>
> All the evidence goes to show that the sun, far from representing only the father, mirrored Schreber's own ambisexuality, being both male and female (p. 378).

The issue of the symbolism of some word or referent would seem one that could be dealt with only by speculation and argument. However, an approach to the matter is possible, which uses internal consistency of the writer's (or speaker's) language as a measure of the similarity of themes. Thus, if Schreber used the same terms in writing of the sun as he did in writing of father figures, this could be taken as evidence that his associations to the two were alike and that in effect they meant the same thing to him. If on the other hand, Schreber wrote about the sun the way he wrote about women, this could be taken as evidence for the Macalpine and Hunter view.

In the present study an effort is made to ascertain the extent of the similarity between the way Schreber wrote about sun and certain other words. The theoretical rationale of the technique to be described has been outlined elsewhere (Laffal, unpublished) and is not requisite in detail to follow the discussion. Briefly, the technique is a development from the viewpoint that meaning may be described as a matrix of associations: verbal associations where one is concerned primarily with verbal meaning and that similarity of meaning may be described in terms of similarity or overlap of associations. This view forms the basis of several studies by Bousfield, Whitmarsh, and Danick (1958). In free speech the contexts of a given word are taken to be the associated matrices of the word, that is, the sets of words which would tend to appear in varying frequencies whenever that word occurred in speech. To make such contextual analyses feasible, it is necessary to categorize the words in the contexts on the basis of synonymity or similarity of reference. The basic criterion for using a category would be the simple one that it be capable of subsuming a fair number of the subject's (or the language's) words.

Many words may be categorized only if one arbitrarily focuses on some special feature of the response. The pertinence of a given word to one category rather than another may be established by definition; however, an analysis of similarity of contexts like the one outlined above would be the ultimate validation criterion to determine if the verbal response belonged in a given class. That is, to the extent that the contextual associates of word x overlapped the contextual associates of other words in Class A, then to that extent, word x would belong in Class A.

---

[1] This report is part of a larger study of Language Distortions in Schizophrenia. The study is supported by United States Public Health Service Grant M-2020.

This paper has been reviewed in the Veterans Administration and is published with the approval of the Chief Medical Director. The statements and conclusions are the result of the author's own study and do not necessarily reflect opinion or policy of the Veterans Administration.

TABLE 1

CATEGORIES OF WORDS IN THE CONTEXTUAL ANALYSIS OF SCHREBER'S LANGUAGE

*agree-*disagree
animal-nature
*begin-*end
*big-little
*body feature-*body function
calm-*upset
clean-*dirty
*come-go
commodity-money
cure-*illness
*durable-*change
*earth-sea
easy-difficult
eat-drink
*essential-trivial
fast-*slow
*forward-*back
free-confine
good-bad
*help-*hurt
*hollow-hill
hot-cold
in-out
*individual-*group
law-crime
lead-follow
*living-dead
male-female
*near-far
new-ancient
open-shut
*play-*work
*possess-*want
*real-unreal
reason-absurd
sacred-profane
*see-*hear
*separate-*join
sex-homosex
*sharp-blurred
*some-all
speak-*write
*strong-weak
*structure-*ruin
true-false
*up-*down
young-mature

Note.—All of the categories in the list were initially applied in the scoring of Schreber's language. However, only the 42 categories identified by asterisks were used in correlating contexts with each other.

A set of 94 categories, grouped roughly into opposition pairs, which had been developed from the study of the speech of a schizophrenic patient, were applied to the analysis of Schreber's autobiography.[2] Heading words for these categories are listed in Table 1, with asterisks showing those categories which were ultimately used in the analysis of the Schreber material. The heading words merely suggest the nature of the words in categories and are not adequately descriptive of the categories.

Unfortunately the third chapter of Schreber's autobiography, which discussed intimately the members of his family, was deleted by censorship, so that there are practically no references to Schreber's father in the autobiography. This makes a direct comparison of speech about father with speech about sun impossible. However, there are some indirect ways of approaching the problem. It is clear in Freud's discussion, that God is the father figure par excellence. Macalpine and Hunter (1955) do not accept this view and consistently with their position about the sun, say: "... his [Freud's] conviction that God and the sun were 'nothing but' father symbols for Schreber; he ignores their obvious female significance..." (p. 378). Some fairly definite predictions arise from these differences of view.

If Freud's contention that sun and God are both father symbols is correct, we ought to find a strong relationship in the way Schreber writes about both of these, and both of them ought to be more strongly related to the way Schreber writes about male figures than the way he writes about female figures.

If the Macalpine and Hunter view is correct, then we ought also to find a strong relationship between the way Schreber writes about sun and God, but both of these ought to be strongly related to the way Schreber writes of females. Their view would seem to entail the prediction that the sun-female and God-female relationship would be at least as strong as, and possibly stronger than the corresponding male relationships.

Two papers by Niederland (1951, 1959) adduce considerable evidence based on a study of Schreber's father's life and works, that the accusations which Schreber directed against God, especially relating to body torture, were also appropriate to Schreber *père*. The older Schreber wrote extensively on the theme that in the earliest years of the child's life a maximum of pressure and coercion should be used in the inculcation of physical and emotional restraint. He devised various mechanical instruments for enforcing correct body posture, not only in waking but during sleep, and he believed in strict disciplinary training measures, including corporal punishment. Niederland (1951) maintains that not only God, but Schreber's physician, Flechsig, who is mentioned often in the autobiography, was a father symbol for Schreber: "First Flechsig, and then God, is reinstated in the place of father" (p. 591). Freud does not make this direct relationship between Flechsig and God and Flechsig and father. Although Freud (1948) does say that "in the patient's mind 'Flechsig' and 'God' belonged to the same class" (p. 49), Freud ultimately identified about God, male, and female. The relationship between Flechsig and male is significantly greater at the .01 level than the relationship between Flechsig and female, Flechsig and God, and God and male. In addition, the relationship between God and Flechsig is significantly greater at the .05 level than the relationship between God and male.

---

[2] The Macalpine and Hunter translation (1955) was the actual text used in the analysis. It was felt that the fine problems of meaning which usually arise in translations would not seriously affect the analysis applied here. This, however, remains unexplored.

DISCUSSION

The findings are rather clear with respect to some of the major hypotheses tested. Several predictions based on Freud's position are supported. Thus, the contexts of God in Schreber's autobiography are significantly more like the contexts of male than of female; the contexts of Flechsig and male are significantly more alike than those of Flechsig and female. The relationship between the contexts of God, male, and Flechsig also suggest that in the slight difference of opinion between Niederland and Freud, the Freudian position may be stronger. Niederland saw both Flechsig and God as father symbols in Schreber's writing. Freud maintained that while both Flechsig and God were in the same class, God was the father substitute, and Flechsig stood in the relationship of a brother. It will be recalled that there are not available in the autobiography enough contexts of father to make direct comparisons possible. In this analysis we have had to settle with contexts of male to test the various hypotheses. Schreber writes about Flechsig more like he writes about male than about God. He writes about God more like he writes about Flechsig than about male. If the matrix of associations around Flechsig is seen as composed of male and father words with male words predominating, and the matrix of associations around God as composed of father and male words, with father words predominating, this could account for the strong relationship between Flechsig and male, the moderately strong relationship between Flechsig and God, and the comparatively weak relationship between God and male.

With respect to the relationship between sun and God as symbols, the findings support neither Freud's nor Macalpine and Hunter's position. Freud saw sun and God in the autobiography as father symbols, and Macalpine and Hunter ascribed feminine qualities to both of them. Unfortunately sun is not written about enough in the autobiography for tests of the reliability of the distribution of its contextual associates, and for enough comparisons with other contexts to provide meaningful tests of significance.

The comparisons made in the present study combine two predictions from both Freud and Macalpine and Hunter: that sun is like God and that both are either like male or like female. These comparisons do not achieve statistical significance, and inspection of the correlations of sun with God suggests that one reason is that sun is not very highly correlated with God.

Table 3 shows female and sun to be more highly related, on the average, than male and sun. Of the four direct comparisons between these two correlations in Table 2, three are in favor of the female-sun correlation, and one in favor of the male-sun correlation. The comparison is statistically inconclusive, but the tendency of the data is in the direction of the prediction derived from Macalpine and Hunter's view that sun is a feminine symbol.

Since the word sun is used relatively infrequently in the autobiography as compared to words like God, Flechsig, male, and female, it is likely that both Freud and Macalpine and Hunter were forced to speculate more freely on its symbolism and relationship to other aspects of Schreber's thinking. In each case some striking context of the word seems to have been taken as the key to its interpretation. Thus, Macalpine and Hunter refer to the fact that Schreber calls the sun a whore. In addition they pointed out that the sun is of feminine gender in German and that there is folk-mythological support for the idea of the sun as an ambisexual thing. Freud laid emphasis on the fact that Schreber himself identified the sun sometimes with the "upper" and sometimes with the "lower" God. Freud also drew on evidence from other cases, from literature, and from mythology to support his assertion that the sun is a father symbol for Schreber. It is possible that Freud and Macalpine and Hunter, in relying heavily on universal interpretations of the sun as a symbol, and having relatively little textual material available in the actual autobiography, failed to evaluate correctly its significance in Schreber's system. It is beyond the scope of this paper to attempt such an analysis, but the findings point to the caution necessary in introducing universal symbolism into the verbalizations of an individual. The complexity of the sun symbolism is well illustrated in a paper by Abraham (1953) who while adhering to the position that the sun is a father symbol, pointed out the many ways in which it is also a female symbol.

The technique that has been used in this study to explore the predictions derived from Freud and Macalpine and Hunter bears some further comment. Much trial and error has gone into the evolution of the method, and it cannot yet be described as a polished procedure. There remain numerous uncertainties about the best way of classifying some words, about whether to broaden or narrow some of the categories, about how many lines of a text should be considered a passage relevant to the key word under study, and so forth. In actually categorizing words, there are many instances in which one's definitions call for a certain categorization of a word but where the categorization seems odd and strained. The procedure itself will undoubtedly undergo numerous transformations as experience accumulates, and such changes may make possible more refined and comprehensive analyses of the same texts. The purpose of the procedure is to provide an objective way of comparing segments of free-flowing language within the theoretical framework that associative similarity is an index of meaning similarity. In the present paper the method has been applied to examining a dispute about the symbolism of certain words in Schreber's autobiography. However, the method is quite general, and it should be applicable wherever problems of similarity of written or spoken stretches of speech arise.

SUMMARY

An objective procedure of contextual analysis was applied to test opposing predictions derived from Freud's (1948) and from Macalpine and Hunter's (1955) studies of the Schreber case. Freud maintained that both God and sun were father symbols for Schreber; Macalpine and Hunter contended that God and sun were ambisexual symbols and stressed their feminine connotations. The contextual analysis involved categorizing words appearing in close association with the key words God, sun, and Flechsig and in close connection with male and female words, and correlating the distributions of categorized responses with each other. It was found that the way Schreber wrote about God was significantly more like the way he wrote about male than the way he wrote about female. Schreber's way of writing about Flechsig was also significantly more like the way he wrote about male than about female. However, it was found that the way Schreber wrote about sun was not similar to the way he wrote about God. Schreber wrote about sun more the way he wrote about female than about male, but this distinction was statistically inconclusive.

REFERENCES

ABRAHAM, K. Restrictions and transformations of scoptophilia in psychoneurotics. (Originally published 1913) In, *Selected papers of Karl Abraham.* New York: Basic Books, 1953. Ch. 9.

BAUMEYER, F. The Schreber case. *Int. J. Psycho-Anal.*, 1956, **37**, 61–74.

BOUSFIELD, W. A., WHITMARSH, G. A., & DANICK, J. J. Partial response identities in verbal generalization. *Psychol. Rep.*, 1958, **4**, 703–713.

FREUD, S. Psychoanalytic notes upon an autobiographical account of a case of paranoia. (Originally published 1911) In, *The complete psychological works of Sigmund Freud.* Vol. XII. London: Hogarth, 1948.

JOHNSON, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.

MACALPINE, IDA, & HUNTER, R. A. Discussion of the Schreber case. In D. P. Schreber, *Memoirs of my nervous illness.* (Trans. by Ida Macalpine & R. A. Hunter) Cambridge: Bentley, 1955.

NIEDERLAND, W. G. Three notes on the Schreber case. *Psychoanal. Quart.*, 1951, **20**, 579–591.

NIEDERLAND, W. G. Schreber: Father and son. *Psychoanal. Quart.*, 1959, **27**, 151–169.

SCHREBER, D. P. *Denkwürdigkeiten eines Nervenkranken.* Leipzig: Mutze, 1903.

SCHREBER, D. P. *Memoirs of my nervous illness.* (Trans. by Ida Macalpine & R. A. Hunter) Cambridge: Bentley, 1955.

(Received November 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 480–481

A CLINICAL NOTE ON THE SPECIFIC NATURE OF AN "ORAL" TRAUMA[1]

BERTRAM P. KARON[2]
Philadelphia Psychiatric Hospital

THE apparent discrepancies between clinical case studies and experimental investigations are disturbing to those of us who believe that there is only one psychology of man, whatever the source of the data.

The importance of oral experiences and traumata in the production of character traits and neurotic and psychotic symptoms has been stressed by Freudian psychoanalysis (Abraham, 1954; Fenichel, 1945, pp. 62–66). The role of these determinants seems to be manifest in case history material, and in many recent modifications of Freudian theory, oral problems are given even greater emphasis (Bergler, 1949; Klein, Heiman, & Money-Kyrle, 1955; Silverberg, 1952). Nonetheless, the bulk of the experimental work, as Orlansky (1949) has noted, does not seem to confirm the importance suggested by theory. It may therefore be worthwhile to discuss the relevant aspects of a patient whose adult symptoms showed clear indications of "oral" pathology but seemed to have no childhood basis for such problems. The nature of his oral trauma may help to clarify the disparity between psychoanalytic reconstructions and experimental findings.

The term "orality" is used in psychoanalysis to refer to the mouth and its functions—especially eating, drinking, sucking, and biting—to the dynamic characteristics connected with these (e.g., dependency is seen as being connected with the experience of being fed), and to the first psychosexual "stage" of early infancy, when these functions of the mouth are assumed to be the most important activities in the life of the child. The term is thus used to refer both to problems which originate in this early period of life and to problems having to do with mouth functions. According to Fenichel, one would expect oral pathology to be prominent in the symptoms of someone who had been excessively deprived of oral gratification, excessively indulged in the oral period, or excessively indulged in the oral period and excessively frustrated in the succeeding stage.

Both in theoretical discussions and in attempts to investigate the importance of "oral" experiences, attention has frequently been focused upon whether or not the child was breast-fed or how early he was weaned. More recently, it has been realized that an abrupt weaning at any age is more traumatic than a gradual one, no matter when it takes place (Sears & Wise, 1950).

Destructive oral experiences have been said many times to play an essential role in the development of schizophrenia (Bellak, 1948, 1958; Fenichel, 1945, pp. 415–452; Rosen, 1953; Sechehaye, 1955). A 32-year-old male chronic paranoid schizophrenic patient who illustrates this contention has been described in a previous paper (Karon & Rosberg, 1958), but the basis for his oral trauma was not spelled out in that brief account.

This patient's delusions were such that any psychoanalytically oriented psychotherapist would have concluded that he suffered severe problems of an obviously oral nature and that he must have undergone an oral trauma of considerable impact. For example, the patient was afraid of being poisoned. He would not drink milk at the beginning of the therapy and confided that: "The Athenian girls are laughing at me. They say their breasts are poisoned."

One day a woman with a rather well-endowed bosom walked by him. He made frantic wiping gestures from his side. Afterwards he said: "She approached me with her breast, and I got a pain in my side."

Despite these symptoms, it was possible to verify that this patient had been breast-fed for a full year and had been gradually weaned. Where, then, could there be an oral trauma?

On one occasion, his mother (who had read some of the literature of psychoanalysis) said to him: "Didn't I give you enough milk? Didn't I give you enough to drink?" The patient stared into space and replied: "The cow gave her calf milk and then kicked it. She shouldn't do that. It's something that happened hundreds of times in the history of the world."

In this delusion of the cow who kicked her calf (the patient, of course, denied that it had anything to do with him) is the key to the oral trauma. It did not consist of being deprived of the breast or of being weaned too early or too abruptly. The oral trauma consisted of a sequence of interactions: The child got fed and then got hurt. The "poisoned" milk seems not to be due to any characteristic of the milk itself, but to the fact that the mother resented any demand upon her to feed the child so that immediately after feeding him, she got angry. The child then felt the impact of her hostility and was hurt. "Poison" is simply something you eat, and you get hurt afterwards.

---

[1] This note was completed during a United States Public Health Fellowship and is based on work previously done.

[2] The writer wishes to express his appreciation to Jack Rosberg, who collaborated in the study of the patient described here, and to John N. Rosen, under whose supervision the patient was treated.

This sequence—child eats, mother gets angry, child feels hurt—was not something which happened once or twice in the patient's life. It was not something that happened to him only during a specific period of childhood. It was something that started when he was an infant and continued all his life. Even at the age of 32, it was observable. The mother took care of the treatment home[3] for a period of time so as to save expenses. She seemed to be making every conscious effort to help her son. The therapist was startled to notice that after every meal she cooked for her son, she had an argument with him. If she did not cook the meal, there was no argument. Thus, it was not merely a single childhood experience, but the continuous re-enactment of the trauma that prevented it from ever being overcome.

It must not be supposed that this was in any way a conscious process for the mother. Consciously, she was doing everything she could for her son. She was unaware of her constant pattern—feed, get mad—which seemed obvious to an observer. Each evening the content of the argument dealt with something else.

Her emotional inability to feed is exemplified by the fact that after raising two children and after many years of married life, her cooking was barely palatable, not only according to the patient but according to the therapist's judgment as well. Her needs were exemplified in the very beginning of the treatment when she complained of her obviously sick son: "He always took and never gave; it's about time he gave a little." Actually, the patient had been forced to give in a number of different ways—his gifts to his mother, his writings, his paintings, etc. (Karon & Rosberg, 1958).

Since it is typical of the schizophrenogenic mother that she cannot tolerate the demand to feed, to take care of, to nurture, it is likely that this repeated pattern of activity—feed, get mad—is frequently present in such cases. The important point is that the oral trauma is not merely a single childhood event, but the continuous re-enactment of the traumatic sequence.

Without being aware of it, the mother continuously imprinted the notion that after eating comes the pain. Is it any wonder then that the patient is afraid of being poisoned?

Thus, the apparent discrepancy between clinical case studies and experimental findings seems to be due to the fact that "oral trauma" has been interpreted as referring to a single specific event—usually weaning. This is not entirely the fault of the investigators since psychoanalytic theory is usually presented as if the key to symptom and character formation lies in such isolated events. The present case suggests that the important "oral" variables are the persisting patterns of interactions, including the mother's involuntary and unconscious reaction to the demand for nurturance. Isolated events, like weaning, are easily influenced by "enlightened" child care education or by conscious good intentions without removing the destructive character of the total relationship. It seems likely that when attention is directed to the continuing relationship between mother and child, the discrepancies between clinical and experimental data will disappear.

REFERENCES

ABRAHAM, K. On the first pre-genital stage of the libido. In, *Selected papers on psychoanalysis*. New York: Basic Books, 1954. Pp. 248–279.

BELLAK, L. *Dementia praecox*. New York: Grune & Stratton, 1948.

BELLAK, L. *Schizophrenia*. New York: Logos, 1958.

BERGLER, E. *The basic neurosis*. New York: Grune & Stratton, 1949.

FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.

KARON, B. P., & ROSBERG, J. Study of the mother-child relationship in a case of paranoid schizophrenia. *Amer. J. Psychother.*, 1958, **12**, 522–533.

KLEIN, MELANIE, HEIMAN, PAULA, & MONEY-KYRLE, R. E. *New directions in psychoanalysis*. London: Tavistock, 1955.

ORLANSKY, H. Infant care and personality. *Psychol. Bull.*, 1949, **46**, 1–48.

ROSEN, J. N. *Direct analysis*. New York: Grune & Stratton, 1953.

SEARS, R. R., & WISE, G. W. Relation of cup-feeding in infancy to thumb-sucking and the oral drive. *Amer. J. Orthopsychiat.*, 1950, **20**, 123–160.

SECHEHAYE, MARGUERITE A. *Symbolic realization*. New York: International Univer. Press, 1955.

SILVERBERG, W. V. *Childhood experience and personal destiny*. New York: Springer, 1952.

(Received February 29, 1960)

---

[3] The patient, while under therapy, lived in a private house with treatment personnel and with his mother.

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 482–483

CRITIQUE AND NOTES

MALE AND FEMALE SUBJECTS IN PERSONALITY RESEARCH

EARL R. CARLSON
Michigan State University

AND

RAE CARLSON
East Lansing, Michigan

There is a sizable accumulation of evidence today that males and females react differently to many social situations and differ in many personality characteristics. Not only are there mean differences on a number of psychological measures, but often a relationship appears for one sex group but not for the other. Potential sex differences are omnipresent, but are largely ignored in even the best of current research on personality and social problems.

The present report is directed to three problems we see in current research: the imbalance in sampling in research, the paucity of tests for sex differences, and the inadequacy of reporting of the sex composition of research samples. An individual investigator, of course, must select his sample to meet specific needs of the problem. Typically, a study is criticized only if the investigator states an obvious overgeneralization. Beyond this, however, it is important at times to view the research literature as a whole to note possible implied overgeneralizations as a function of consistent biases entering into the selection of samples. From such views we can often redirect our efforts, and explicitly check for the limits of generalizability of early findings.

The 298 empirical studies reported in the 14 consecutive issues of the *Journal of Abnormal and Social Psychology*, Vol. 56 (1958), Number 1 through Vol. 60 (1960), Number 2, were selected as a sample of current research on personality and social phenomena. The few studies based upon single cases, nonhuman subjects, or reanalyses of previous research data were not included in this sample. Tabulations were made of the sex composition of the samples of subjects studied, the completeness of reporting of sex composition, and the reports of tests for, and findings of, sex differences.

The results show an overwhelming use of males as compared with female subjects. Of a total of 298 studies reported, 113 (38%) used male subjects exclusively, and 15 studies (5%) used only female subjects. A total of 108 studies reported using both male and female subjects, and for 65 studies the sex of subjects was unknown. The studies using both male and female subjects were fairly well balanced; in 36 studies (33%) an equal number of males and females were used, and in 20 studies (19%) the proportions of males and females were not reported.

Obviously, sex differences could be tested in only those 108 studies using subjects of both sexes. In only 32 studies (30%) were tests of sex differences reported. But in 22 of these cases (69%) sex differences were found at a level of statistical significance.

Several implications follow from these findings. First, males are selected far more often than females as research subjects. In studies of many psychological processes this bias may lead to no problems, but on a number of personality and social measurements sex differences do occur. As readers of research reports we can say to ourselves that of course we realize that only males were used, and that generalizations to females must depend upon further research. It is not only difficult to retain such contingencies in our recollection of studies we have read, but, we venture to guess, readers are largely unaware of sex composition in reported studies—even when this information *is* reported.

How often are sex differences actually tested? If it is correct that all *tests* for differences are reported, then with 69% of these showing significant differences it would follow, with little question, that differences do in fact exist in numerous cases, but were overlooked by the researcher. If many tests were made but never reported, then the reported differences could be expected by chance alone. Either of these alternatives would seem more plausible than the possibility that unreported tests precisely parallel the sex differences reported.

Even for a reader who is cautious in generalizing results, interpretation is ambiguous when the sex composition of the sample is incompletely reported. In 20 of 108 studies reporting both male and female subjects the proportions are unknown. In 62 studies (21% of all studies) a reader is totally unaware of the sex of the subjects, although many of these studies presumably included both men and women. In several studies the *N* can be discerned only from a thorough reading of the tables. We suspect that all readers will assume that Yale undergraduates and naval recruits are male, as we have done in the tabulations for this paper. There may be other similar situations where the authors assume, perhaps too optimistically, that readers can infer the sex of subjects from other information given.

Mean sex differences may or may not affect tests of psychological relationships between variables. In cases where test ceilings restrict the range of scores for one group, the variance for this group may be reduced to the point where a relationship cannot be tested. In a substantial number of studies a relationship holds for one sex but not for the other. A more serious problem arises in studies using extreme groups, where mean differences between sexes may lead to an imbalance of sex in the groups selected for intensive analysis. Some studies separate sex distributions and select high and low groups separately from each (e.g., Goldstein, 1959). Other studies do not (e.g., Taylor, 1958) and the resulting findings may stem from other correlates of sex than the variable under study.

Obviously many characteristics of an individual may interact with psychological relationships under study. A researcher cannot hope to do more than control or test for the effects of more than the most significant of these factors. We suggest that for two reasons, at least, sex should be included routinely as an explicit variable in studies where such a comparison is possible. First, sex differences do run rampant through our literature, and no one can deny that they are often important. Secondly, sex is an easy factor to measure, and sex differences are easily tested and reported. There seems little excuse not to include them as a standard procedure.

REFERENCES

GOLDSTEIN, M. J. The relationship between coping and avoiding behavior and response to fear-arousing propaganda. *J. abnorm. soc. Psychol.*, 1959, **58,** 247–252.

TAYLOR, J. The effects of anxiety level and psychological stress on verbal learning. *J. abnorm. soc. Psychol.*, 1958, **57,** 55–60.

(Received May 2, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 484-485

PERSONALITY SIMILARITY, POSITIVE AFFECT, AND INTERPERSONAL ATTRACTION[1]

CARROLL E. IZARD
Vanderbilt University

RECENT studies have shown that perceived similarity (Newcomb, 1956), assumed similarity (Fiedler, 1958), and actual similarity (Izard, 1960) are significant correlates of interpersonal attraction. The latter paper postulated that interpersonal positive affect—defined as the expression of favorable feeling, self-involving interest, acceptance or esteem in relation to another person—was a significant factor in much interpersonal behavior and a key determinant of interpersonal attraction. The role of actual personality similarity was assumed to be that of facilitating the mutual expression of positive affect. The paper argued that persons with similar affective characteristics tend to find the same interpersonal situations satisfying and that they realize more accurate interpersonal perception and communication. The finding of the previous study that there was significant personality profile similarity between pairs of mutual friends was considered as support for these arguments.

Like the previous study, the present one assumed that interpersonal positive affect is a key determinant of interpersonal dynamics and that personality similarity facilitates the mutual expression of positive affect. The hypothesis was that subjects and their unilateral sociometric choices have significantly similar personality profiles prior to acquaintance, while subjects and their sociometric rejections do not.

METHOD AND RESULTS

Edwards' (1954) Personal Preference Schedule (EPPS), the personality measure utilized in assessing profile similarity, was administered to an entire freshman class on entering college. Six months later 47 of the female students in this group were given a sociometric form requiring them to list the three most likeable and the three least likeable girls in their class. The instructions emphasized that each subject should list the three whom she personally felt were most likeable and should avoid listing people on the basis of general popularity. They were requested to exclude from consideration anyone whom they knew prior to entering college.

Twenty-five of the subjects made independent (mutually exclusive) sociometric choices. Twelve of the 25 were randomly selected for pairing with classmates they listed as most likeable and the 13 remaining subjects were paired with the classmates they listed as least likeable. As shown in Table 1, the analysis of variance technique for comparing profiles of pairs (Izard, 1960) indicated that prior-to-acquaintance personality profiles were significantly similar for subjects and their sociometric choices ("most likeable" classmates) but not for subjects and their sociometric rejections ("least likeable" classmates).

TABLE 1
ANALYSIS OF VARIANCE COMPARING PERSONALITY PROFILES OF PAIRS BASED ON SOCIOMETRIC CHOICES OF MOST LIKEABLE AND LEAST LIKEABLE CLASSMATES

Source	*df*	*MS*	*F*
Between Variables	14	475.75	
Variables × Groups	14	15.98	
Variables × Pairs (Sociometric choices)	154	22.25	1.45**
Variables × Pairs (Sociometric rejections)	168	16.96	
Variables × Subjects (Sociometric choices)	168	15.38	
Variables × Subjects (Sociometric rejections)	182	25.27	

** $p < .01$.

DISCUSSION

Izard's (1960) earlier study showed that personality profile similarity was a significant correlate of the mutual friend relationship. That study did not show whether the personality similarity existed prior to the friendship or was a result of it.

The present study demonstrated that actual personality similarity was an antecedent of unilateral sociometric choice (nomination as most likeable) but not of sociometric rejection (nomination as least likeable). This finding lends further support to the assumption that personality similarity facilitates the expression of interpersonal positive affect. One aspect of interpersonal positive affect consists of favorable feeling or liking for another individual, precisely the factor that was measured by the sociometric form employed in this study.

---

[1] This study was supported by Vanderbilt-ONR Contract Nonr 2149(03), C. E. Izard, Principal Investigator. The opinions and conclusions do not necessarily reflect those of the United States Department of the Navy.

REFERENCES

EDWARDS, A. L. *Manual, Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

FIEDLER, F. E. *Leader attitudes and group effectiveness.* Urbana: Univer. Illinois Press, 1958.

IZARD, C. E. Personality similarity and friendship. *J. abnorm. soc. Psychol.*, 1960, **61**, 47–51.

NEWCOMB, T. M. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–586.

(Early publication received July 6, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 485–487

THE STABILITY OF COGNITIVE CONTROLS[1]

RILEY W. GARDNER AND ROBERT I. LONG
Menninger Foundation

In a recent monograph, Gardner, Holzman, Klein, Linton, and Spence (1959) developed a conception of dimensions of cognitive control and described a first study of relations between previously defined control principles. However, no adequate studies of the reliability of measures of cognitive controls have yet been published.

A recent study provided an opportunity to retest a group of subjects who had been given criterion tests for several control principles approximately 3 years earlier (mean interval, 36.6 months). The cognitive controls involved (see Gardner et al., 1959) are: Equivalence Range, Scanning, Leveling-Sharpening, and Constricted-Flexible Control.

METHOD

At the end of 3 years following the initial testing of a group of 80 female subjects, 38 were available for retesting. At the time of the initial testing, these 38 subjects were students at a small midwestern university and ranged in age from 18 to 22, mean age, 19.3.

The control principles involved and the tests and scores used to represent them are summarized below.

Equivalence Range. Object Sorting Test: An exact replication was performed. The score is the number of groups formed by the individual in free sorting of 73 objects.

Scanning. Size Estimation Test I: The original size estimation score used to represent this control principle was a presumed correlate of the subject's scanning strategy: average error in judging the sizes of plain circular disks. Results of subsequent studies of the Scanning principle (Gardner, 1959, in press) supported the impression of Gardner et al. (1959), however, that constant error is a more appropriate measure than average error to represent the effects of scanning on relative apparent magnitude. An exact replication of judgments of three 40-millimeter disks (gray, 7 gm.; black, 7 gm.; heavy gray, 67 gm.) was performed. In this test, the subject held the disks in her left hand while she adjusted a variable circle of light to be equal to it in size. Scores were constant error in judging each disk and constant error for the entire test.

Leveling-Sharpening. Schematizing Test: An exact replication of this test was performed. It should be noted, however, that this test is not repeatable in the same sense as the others discussed here. Effective measurement of assimilative interaction among percepts and memory traces in this test—which involves 150 size judgments of squares ranging from 1.3 to 13.7 inches—requires that the subject be naïve as to the range of sizes and the sequence of presentation. The gradual progression of stimulus sizes (over 35 minutes) is from smaller to larger. A subject who has experienced the full range of sizes approaches retesting with considerable information concerning the general structure of the test. Some subjects have an "aha" experience late in initial testing, when they realize suddenly that they are lagging far behind the trend of size increase. Such subjects abruptly adjust their judgments upward, at times with accompanying written or vocal comments. Other subjects remain unaware of their lagging throughout, even though they may judge the 13.7-inch square as "five inches." Scores used were percentage ranking accuracy, mean percentage increment error (see Gardner et al., 1959), and a recently developed "lag" score, which consists of the slope defined by the logarithms of the actual means of the last nine series of 15 judgments minus the slope of the logarithms of the subject's judgments. This new score omits judgments in the first series and is based equally on the remaining series means. The increment error score is based on the mean percentage increment of the last nine series means over the mean of the first series and hence is overly dependent on the mean of the first series of judgments.

Constricted-Flexible Control. Color-Word Test: Exact replication of all three parts of this test (reading color names, color strips, and colors printed in incongruous color names) was performed. Scores consisted of the reading times for each part and the interference score: the difference between the subject's time on Part III and the time predicted on the basis of the regression of time for Part III on time for Part II.

RESULTS

Means and standard deviations for all tests and measures are presented in Table 1. Means of all constant error scores for Size Estimation Test I are significantly lower on retest ($p < .01$), as

---

[1] This investigation was supported by Research Grant M-2454, from the National Institute of Mental Health, United States Public Health Service.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF TEST AND RETEST SCORES

Score	N	Mean		Standard Deviation	
		Test	Retest	Test	Retest
Object Sorting Test					
Number of Groups	34	17.1	15.7	7.8	6.8
Size Estimation Test I					
Constant Error, Gray Disk	33	2.85	2.02	1.75	1.43
Constant Error, Black Disk	33	3.15	2.15	2.22	1.50
Constant Error, Heavy Gray Disk	33	2.95	2.02	2.04	1.71
Constant Error, Total Test	33	2.96	2.06	1.87	1.46
Schematizing Test					
Percentage Accuracy	38	66.87	64.21	14.75	13.16
Mean Percentage Increment Error	38	73.39	67.31	37.19	38.89
Lag	38	.0158	.0153	.0215	.0192
Color-Word Test					
Reading Time, Color Names	33	41.4	41.7	6.4	6.3
Reading Time, Colors Alone	33	56.6	56.4	7.1	6.8
Reading Time, Color-Word	33	97.4	90.5	12.2	11.5

evaluated by t tests for correlated measures. None of the other test-retest means are significantly different.

Pearson correlations between the test and retest scores are presented in Table 2.

Additional findings bearing upon the reliability of these measures include the following:

In the original testing, r's between constant error scores for the three disks used in Size Estimation Test I were: gray:black, .81; gray:heavy gray, .80; black:heavy gray, .82. In the retest, these r's were .90, .80, and .82, respectively.

In the original testing, the r between the accuracy and mean percentage increment error scores for the Schematizing Test was −.22. In the retest, it was −.37. The r's between the accuracy and lag scores were −.26 and −.18 for the two testings.

In the original testing, the r between reading time for the colors-alone and color-word sections of the Color-Word Test (used to obtain the interference score) was .64. In the retest, this r was .61.

TABLE 2

PEARSON CORRELATIONS BETWEEN TEST AND RETEST SCORES

Score	N	r
Object Sorting Test		
Number of Groups	34	.75***
Size Estimation Test I		
Constant Error, Gray Disk	33	.64***
Constant Error, Black Disk	33	.57***
Constant Error, Heavy Gray Disk	33	.63***
Constant Error, Total Test	33	.68***
Schematizing Test		
Percentage Accuracy	38	.52***
Mean Percentage Increment Error	38	.36*
Lag	38	.45**
Color-Word Test		
Reading Time, Color Names	33	.54***
Reading Time, Colors Alone	33	.75***
Reading Time, Color-Word	33	.75***
Interference	33	.55***

* $p < .05$.
** $p < .01$.
*** $p < .001$.

DISCUSSION

The test-retest correlations observed for this group of subjects offer support for the assumption that cognitive controls, as measured by these procedures, are relatively enduring features of cognitive organization. The conception of "low rate of change" processes does not imply exceptionally high reliability, particularly in the present group of relatively young subjects. All the subjects involved in this study underwent major changes of life situation during the interval: 19 of the total N of 38 had been graduated from their university, all but 13 were married, and 4 were pregnant at the time of retesting. All these factors could have contributed to a reduction of r's between test and retest scores.

Except for the significant reduction of constant errors in Size Estimation Test I—which may imply an increase in extensiveness of scanning in the group as a whole (Gardner, 1959; Gardner et al., 1959)—none of the retest means is significantly different from that in the original testing. Apparently, the cognitive controls involved in response to three of the four tests were relatively stabilized at the time of the initial testing.

The lower reliability coefficient for the interference score of the Color-Word Test than for reading times on the colors-alone and color-word sections of the task is attributable to the fact that reliability of the interference score is a function of the reliabilities of both scores used to compute it, plus the correlation between them.

As anticipated, test-retest correlations are lowest for the Schematizing Test scores. This does not necessarily imply that individual differences in assimilative interaction among percepts and traces are more variable over time than the other cognitive functions sampled. As noted earlier, some subjects, more than others, become aware of the actual range during initial testing. The somewhat higher *r* for the lag score suggests that it may provide a more effective estimate of assimilative interaction than the earlier mean percentage increment error score.

REFERENCES

GARDNER, R. W. Cognitive control principles and perceptual behavior. *Bull. Menninger Clin.*, 1959 **23**, 241–248.

GARDNER, R. W. Cognitive controls of attention deployment as determinants of visual illusions. *J. abnorm. soc. Psychol.*, in press.

GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, HARRIET B., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Issues*, 1959, 1, No. 4.

(Early publication received July 18, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 487–488

PRESTIGE EFFECTS IN BODY-SWAY SUGGESTIBILITY

J. P. DAS

Ravenshaw College, Cuttack, India[1]

THIS study is concerned with the effect of the prestige of the experimenter (*E*) on the body-sway of the subject (*S*). Two orthogonal factors of suggestibility, primary and secondary, have been distinguished by Eysenck (1947), who classifies body-sway suggestibility under primary suggestibility.

METHOD

Four prestige levels were introduced. Three of these represented three hierarchical levels of prestige, the fourth was of neutral prestige value. The body-sway suggestions of the four *E*s representing the four prestige levels were tape recorded. Their manner of reading out the suggestion was made as similar as possible, the loudness of their suggestions was kept more or less equivalent by the volume adjuster during playback. The four *E*s were: I. Head, Psychology Department, II. Laboratory Assistant, III. Attendant of the Department, and IV. Unidentified. Before each one spoke, the author's voice introduced the speaker giving his name and designation. The unidentified one was introduced as: "Now you are going to hear some one speak; who it is I don't know, for the recording was done when I was absent."

The four prestige levels (I, II, III, and IV) presented in a sequence, admit of 24 permutations. Two *S*s were taken for each permutation, thus requiring 48 *S*s. Static ataxia of *S* for 60 sec. was measured before body-sway suggestions were introduced. Brief intervals not exceeding 60 sec. occurred between the termination of one suggestion and the beginning of the next. Suggestions of any *E* lasted for 2.5 min. During these intervals *S*s sat down in a nearby chair. *S*s were tested individually. The maximum forward sway under such prestige level was recorded.

Forty-eight male college students of the Psychology Department, aged 16–23, were used as *S*s. All *S*s were new to the experiment. All of them knew the Head of the Department, Laboratory Assistant, and the Attendant personally.

To measure body-sway, two sliding pieces of small aluminum plates were mounted on a ruler. Through the sliding plates passed a fine thread. To the part of the thread between the two plates was fixed a tiny piece of metal which could be moved manually, but would not move by striking the metal plates. The thread passed through a pulley at each end of the ruler which was fixed to the wall in a horizontal position. Attached to one end of the thread was a little weight while the other end was to be attached to *S*'s collar by means of a safety pin. When *S* swayed, the tiny metal would strike the piece of aluminum plate either to the left or to the right side according to whether *S* swayed forward or backward. The ruler was 40 in. long, so the aluminum plates were brought to the twentieth in. which served as a central point before sway was recorded. Maximum forward and backward sways could be measured easily by noting the last positions of the left and right side plates immediately after the suggestions were over.

RESULTS AND DISCUSSION

The means and standard deviations for amount of sway under the suggestions of the four *E*s differing in prestige are shown in Table 1. A fall was scored as 7 in., for beyond that a person would always lose his balance. An analysis of variance was performed taking into account the four prestige and the four order effects. The order effects refer to whether any prestige level was presented

---

[1] Full address: Postgraduate Department of Psychology, Ravenshaw College, Utkal University, Cuttack 3, India.

TABLE 1

BODY-SWAY (IN INCHES) UNDER DIFFERENT PRESTIGE LEVELS

($N = 48$)

	Levels of prestige			
	I	II	III	IV
Mean	2.21	1.90	1.75	1.66
SD	2.96	2.70	2.49	2.37

TABLE 2

ANALYSIS OF VARIANCE FOR BODY-SWAY SCORES UNDER HIGH AND LOW PRESTIGE CONDITIONS

Source	*SS*	*df*	*F*
Total	667.716	191	
Between	22.277	1	6.558*
Within	645.439	190	

* $p < .05$.

in the first, second, third, or fourth place. For instance, Prestige Level I was presented first to 12 *S*s and similarly in the second, third, and fourth places to different sets of 12 *S*s. However, neither the main effects nor the interaction was significant. Since the Laboratory Assistant, Attendant, and Unidentified person together can be regarded as *E*s of low prestige value in comparison with the Head of the Department, it was thought justified to compare Levels II, III, and IV combined with Level I. Accordingly a simple analysis of variance was made (Table 2) neglecting the order effect, and the F was significant ($p < .05$).

In order to see the differential effect of prestige on relatively suggestible and nonsuggestible *S*s their respective scores were further analysed. The first ten high- and low-scoring *S*s were chosen on the basis of their total suggestibility scores under the four prestige levels. It was predicted that the high suggestibility group would show the differential effect of prestige, whereas the low suggestibility group would not do so. The sign test (Siegel, 1956) was used on the scores of these 20 *S*s. Comparisons were made only between prestige Levels I and II, I and III, and I and IV, since their differences alone had been previously found to be significant in an earlier sign test. None of these differences were significant for the low-suggestibility group. The difference between I and IV alone was significant ($p = .011$, one-tailed test) in the high-suggestibility group.

The main findings may be summarized as follows:

1. Prestige increases body-sway suggestibility.
2. *S*s high in suggestibility, unlike those low in suggestibility, showed a significant increase in body-sway under prestige.

REFERENCES

EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received July 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 489–490

SOCIAL CONFORMITY IN PERCEPTION OF THE AUTOKINETIC EFFECT[1]

NORMAN S. ENDLER[2]

Division of Counseling, Pennsylvania State University

Sherif (1935) used the autokinetic (AK) technique[3] to study the development and maintenance of social norms, and concluded that social norms evolve in group situations and may subsequently manifest themselves when the subject (*S*) is alone. Since the *S*s use subjective scales in developing AK norms, the experimenter (*E*) cannot objectively control the norms.

Luchins (1954) modified Sherif's technique by varying the illumination surrounding the AK light. He and Paul (1950) have shown that the point of light becomes fairly stable over trials for an *S* after the surrounding illumination reaches a certain intensity. Thus, the amount of illumination needed to stabilize the point of light (for an *S*), the point of cessation (of AK movement), is a reliable measure of the threshold below which the *S* observes the AK movement. Three advantages of his method are: it introduces stimulus evidence, it is objective, and the social norms can be controlled by *E*.

In this study Luchins' technique was used to test two hypotheses: (*a*) the AK point of cessation can be raised or lowered by special conformity pressures; and (*b*) the socially induced changes are relatively permanent, i.e., true conformity rather than compliance occurs.[4]

METHOD

The AK apparatus was a black light-proof wooden box, 12″ × 12″ × 14″, with a 4″-square aperture in the center of its front face. The box was divided by a vertical black cardboard with a 2.8 mm. center hole. A 40-w. bulb behind the partition (the AK stimulus) burned continuously. A 100-w. bulb connected to a variac, placed in front of the partition served as the variable objective field surrounding the AK light. Twenty grade school and high school students were randomly divided into two groups: "stopped moving" group (9 boys, 1 girl); and "always moving" group (7 boys, 3 girls). *S*s were tested under four different sequential conditions or treatments.

First, a pretreatment reference point was established. *S*, seated 12 ft. from the front of the AK box, was told to report everything he saw. After *S* first reported movement of the AK light, *E*, seated behind *S*, gradually increased the surrounding illumination from zero variac units until *S* reported cessation of the AK movement. The variac reading at this point, when converted to watts, was an index of the point of cessation. The illumination was then increased to a maximum of 119 w. (ascending series). For the descending series the illumination was decreased from 119 w. until *S* reported AK movement again. This point also served as an index of the point of cessation. Three ascending and three descending series were averaged separately[5].

One day after the pretest, *S* was tested in the presence of an accomplice. For the Stopped Moving group, the experimental aim was to lower the *S*'s point of cessation. Pretest procedures were modified as follows: The accomplice always responded before *S*. On *E*'s signal the accomplice began reporting cessation of movement at a level of illumination well below the point where *S* usually reported cessation and continued until maximum illumination was achieved. For the Always Moving group, the aim was to raise *S*'s point of cessation. The procedure was the same as for the Stopped Moving group except that the accomplice always reported AK movement irrespective of illumination.

Posttests I and II were the same for both groups and were exactly the same as the pretest. Posttest I occurred immediately after the experimental treatment. Posttest II occurred two weeks later.

The points of cessation were averaged,[6] producing one ascending and one descending threshold score (in watts) for each *S* for every treatment.

RESULTS AND DISCUSSION

Means and standard deviations are presented in Table 1. Four analyses of variance of the two-factor type were computed, and showed in each case that differences among session means were significant ($p < .001$).[7]

---

[1] This paper is based on a master's thesis conducted at McGill University in 1954. The author is indebted to A. S. Luchins who directed this thesis, for his helpful comments and suggestions. A modified form of this paper was presented at the Illinois Psychological Association in April 1957 in Peoria, Illinois. The author also wishes to thank Donald H. Ford for his aid with the final draft.

[2] Now at York University, Toronto, Canada.

[3] The perception of movement of a stationary pinpoint of light, fixated in a completely dark room, is called the autokinetic effect. This effect can be induced in most *S*s and creates a situation which is new to most *S*s.

[4] For this study the empirical basis suggested by Festinger (1953) for distinguishing between true conformity and compliance was used. True conformity implies that changes due to social pressure are maintained in a subsequent nonsocial situation, and compliance implies that socially induced changes are not maintained in subsequent nonsocial situations.

[5] The ascending and descending series were analyzed separately because *S*s often complained about afterimages in the descending series.

[6] See Footnote 5.

[7] For the Always Moving group there were also sig-

TABLE 1

MEANS AND STANDARD DEVIATIONS (IN WATTS) FOR POINTS OF CESSATION OF AUTOKINETIC MOVEMENT

Group	Series	Sessions			
		Pretest	Accomplice Present	Immediate Posttest	Two-week Posttest
Stopped Moving	Ascending				
	M	12.76	7.82	8.30	7.47
	SD	1.18	0.65	1.05	1.15
	Descending				
	M	11.24	7.52	8.30	6.04
	SD	2.05	0.60	1.05	2.39
Always Moving	Ascending				
	M	11.37	15.31	15.69	16.22
	SD	1.59	3.21	3.74	4.14
	Descending				
	M	10.48	16.46	15.69	14.56
	SD	1.36	4.29	3.74	3.27

Tukey's (1949) tests for comparing individual means in the analysis of variance revealed that the pretest means differed significantly ($p < .01$) from the treatment and posttest means. For the Stopped Moving group the treatment and posttest means were significantly larger ($p < .01$) than the pretest means. For the Always Moving group the opposite was true.

The results indicate that social pressure produced a significant change in the point of cessation of AK movement (a relatively stable norm). The point of cessation was regularly either raised or lowered depending on the nature of the social pressure and was in the direction of conformity to an accomplice's judgments. The new point of cessation was relatively stable, and manifested itself in nonsocial situations both immediately after, and two weeks after the *S*s were subjected to social pressure. Thus, apparently true conformity rather than public compliance occurred in the social pressure situation.

REFERENCES

ENDLER, N. S. The influence of social factors on the autokinetic effect. Unpublished master's thesis, McGill University, 1954.

FESTINGER, L. An analysis of compliant behavior. In M. Sherif and M. O. Wilson (Eds.), *Group relations at the crossroads.* New York: Harper, 1953.

LUCHINS, A. S. The autokinetic effect and gradations of illumination of the visual field. *J. gen. Psychol.*, 1954, **50**, 29–37.

PAUL, I. H. The modified autokinetic technique. Unpublished master's thesis, McGill University, 1950.

SHERIF, M. A study of some social factors in perception. *Arch. Psychol., NY*, 1935, **27**, 187.

TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.

(Received July 28, 1959)

---

nificant differences among *S*s. Individual differences, however, were not of prime relevance to this study and therefore there were no further analyses of the differences among *S*s.

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 491–494

A FACTOR ANALYSIS OF THE WESLEY RIGIDITY INVENTORY:

ITS RELATIONSHIP TO AGE AND NONVERBAL INTELLIGENCE

SHEILA M. CHOWN
Medical Research Council, University of Liverpool

THE Wesley Rigidity Inventory has been previously described in this journal by Wesley (1953). It consists of statements rated by five psychologists as indicating a high degree of rigidity in those who agree with them. Total score on the inventory has been found to correlate to a significant extent with other tests said to measure rigidity. For instance, it correlated with certain aspects of concept formation in a card sorting test (Wesley, 1953), with behavior in a group level of aspiration test (Zelen & Levitt, 1953), and with score on the California F Scale (Jones, 1954).

In the original form 50 items were used; later a 41-item version of the inventory was developed. Zelen and Levitt (1954) pointed out that a casual reading of the 41-item version suggested an overlap between a number of items. By inspection of the original statements, these authors chose 12 which they considered did not overlap, and proceeded to show that this form of the inventory was as reliable as the 41-item version, that the distribution of scores obtained with the two versions was similar, and that similar levels of correlation were obtained between the long and short forms and the California E Scale. However no factor analysis or item analysis appears to be available, so that there is no evidence to suggest that either form of the test is unidimensional. The object of the present analysis was to find out whether the inventory is unidimensional, and if not, to ascertain the relationship between age of subjects, nonverbal intelligence, and the separate dimensions.

METHOD

Procedure and Subjects. The Wesley inventory (39 statements from the 41-item version) was given as part of a battery of 15 paper-and-pencil tests designed to investigate various aspects of rigidity in relation to chronological age. The battery, selected from reports of previous work on rigidity (Chown, 1959), also included Raven's Progressive Matrices (1938) and the Mill Hill Vocabulary (synonyms) Scale (Raven, 1943).

The battery was completed in group testing sessions by 200 male subjects aged between 20 and 82. Subjects were obtained from military hospitals, foreman courses, veterans clubs, Worker's Educational Association classes, and "over 60's" clubs.

Table 1 shows the age distribution of the sample. The distribution of Progressive Matrices scores was approximately normal, with a mean of 40.2 and a standard deviation of 9.1. On this measure of nonverbal intelligence the scores decreased slightly from 20 to 60 years and then more rapidly. The Vocabulary scores (mean 20.9 and standard deviation 6.0) tended to rise with age until 60 and then fall off.

Analysis. The true-false responses to the 39 items, chronological age, nonverbal intelligence, and vocabulary score for the 200 subjects were intercorrelated. Tetrachoric coefficients were used between items; biserial *r*'s for items with other tests; and product-moment *r*'s between age, nonverbal intelligence, and vocabulary. The distribution of all intercorrelations is shown in Table 2. The median correlation was 0.13.

TABLE 1

Age	Number of Subjects
20–29	54
30–39	50
40–49	33
50–59	39
60+	24

TABLE 2
DISTRIBUTION OF INTERCORRELATIONS IN ORIGINAL MATRIX

Number of correlations	Size of correlations	p
456	0.00–0.13	>.05
98	0.14–0.17	.05–.01
81	0.18–0.22	.01–.001
226	0.23–0.74	<.001

The matrix was factor analyzed by the centroid method. McNemar's (1942) criterion was used to test for completeness of factor extraction, and factoring was stopped after the third factor had been obtained. Seven orthogonal rotations were carried out. The three factors together accounted for 24.1% of the variance.

RESULTS

The items marked with an asterisk in Table 3 have been selected to form factor scores. Inspection of the loading patterns suggested that good representation of each factor could be achieved without serious loss of purity by adapting a general criterion loading of 0.24 for the inclusion or rejection of an item. (There is no generally accepted test of significance for a factor loading, but the figure chosen in this instance is close to that for twice the *SE* of a zero r_{tet} for $N = 200$: viz., 0.22.)

TABLE 3

Rotated Factor Loadings

Item number	I	II	III
1	*29	04	−01
2	04	*48	−04
3	*48	00	−17
4	11	*28	−19
5	−02	*49	11
6	−09	*42	22
7	10	14	*52
8	09	*43	−08
9	20	16	*26
10	14	03	17
11	−29	30	20
12	50	28	16
13	−17	*40	22
14	−38	26	36
15	*39	−10	18
16	13	*36	−04
17	−24	23	*34
18	−17	20	*43
19	*56	16	12
20	*63	16	−24
21	07	39	30
22	*59	14	10
23	19	*66	−10
24	42	25	−08
25	−04	20	−12
26	36	25	−05
27	12	19	−08
28	16	*42	24
29	04	*30	−12
30	10	−10	−17
31	15	14	*33
32	*34	16	10
33	40	−17	39
34	−12	*53	00
35	28	40	−39
36	*41	18	18
37	*60	06	18
38	*36	14	20
39	10	06	*33
Raven	−04	*−34	−20
Vocab.	08	−38	56
Age	14	12	*63

Twenty-nine items were loaded above 0.24 on one of the factors and below 0.24 on the other two. There were 10 such items in Factor I, 12 in Factor II, and 7 in Factor III. The items are recorded in Tables 4, 5, and 6, where it can be seen that age loaded highly on Factor III and reversed Raven score or "unintelligence" on Factor II. Thirteen items were not counted in any factor, four because they were low on all, and nine (including Vocabulary score) because they were high on more than one.

The selected items, when summed with unit weighting, predict their respective factors 0.84, 0.90, and 0.68. Intercorrelations between the scores are as follows: I and II, 0.23; I and III, 0.17; II and III, 0.21. While these coefficients are all significant below the .01 level, their small size indicates that the three scores are not highly related to each other.

Discussion

Of the three factors, II is the easiest to identify. Both nonverbal intelligence and vocabulary have a high negative loading on it (see Table 3). Inventory items included in it consist of liking for slow steady methods, dealing with one thing at a time, sticking to own ways or beliefs, and statements about "people-in-general." It is not surprising to find that this type of rigidity is associated with low intelligence.

Factor III, with age having the highest loading, does not include many inventory items. The six which do load highly, however, present a consistent picture of liking for established routine.

Factor I, not associated with either intelligence or age, is made up of items dealing with persistence, methodical behavior, and liking for detail.

In view of the high correlation between age and nonverbal intelligence it had been thought that the analysis would produce one factor on which both age and intelligence would load highly, and perhaps another factor, not associated with either. The interesting thing about these results is that age and nonverbal intelligence do *not* load highly on the same factor. One group of items is related to chronological age and another to this measure of intelligence. The third group of items, as expected, is not related to either intelligence or age.

It is possible to examine the composition of Zelen and Levitt's short form of the inventory in the light of the present results. When describing their method of selection, they gave an example of two items which they thought were opposites and another example of two items which they thought so similar as to overlap. Their "opposite" items (present Numbers 12 and 31) do not load highly on the same factors. Moreover, their "overlapping" items (present Numbers 19 and 23) appear one in Factor I and the other in Factor II. These pairs are, it seems, not comparable after all, and presumably some information is lost by eliminating one of them. When their 12 final items are examined, it is found that 4 of these do not appear in the version used here, but 8 are common to the short and long versions. Five of the eight items are present in the factors, three loading highly on I, one on II, and one on III. Of the three remaining items, one loads highly on all three factors, one on I and II, and one on I and III. Thus it seems that consistent with the general contention of Zelen and Levitt the short form of the inventory approximates in factor make-up to the structure of the main inventory, though Factor II is underrepre-

TABLE 4

SCORED ITEMS FOR FACTOR I

Item number	Loading	Item
20	0.63	I am usually able to keep at a job longer than most people.
37	0.60	I am (not) inclined to go from one activity to another without continuing with any one for too long a time.
22	0.59	I always finish tasks I start, even if they are not very important.
19	0.56	I am a methodical person in whatever I do.
3	0.48	I prefer work that requires a great deal of attention to detail.
36	0.41	I usually dislike to set aside a task that I have undertaken until it is finished.
15	0.39	My interests (do not) tend to change quickly.
38	0.36	I prefer to do things according to a routine which I plan myself.
32	0.34	I believe that promptness is a very important personality characteristic.
1	0.29	I am often the last one to give up trying to do a thing.

TABLE 5

SCORED ITEMS FOR FACTOR II

Item number	Loading	Item
23	0.66	People who go about their work methodically are almost always the most successful.
34	0.53	It is the slow steady worker who usually accomplishes the most in the end.
5	0.49	I prefer doing one thing at a time to keeping several projects going.
2	0.48	There is usually only one best way to solve most problems.
8	0.43	I usually maintain my own opinions even though many other people may have a different point of view.
6	0.42	I dislike to change my plans in the midst of an undertaking.
28	0.42	I have never done anything dangerous for the thrill of it.
13	0.40	I would not like the kind of work which involves a large number of different activities.
16	0.36	I usually find that my own way of attacking a problem is best, even though it does not always seem to work in the beginning.
Raven	−0.34	Nonverbal intelligence
29	0.30	It is always a good thing to be frank.
4	0.28	I often become so wrapped up in something I am doing that I find it difficult to turn my attention to other matters.

TABLE 6

SCORED ITEMS FOR FACTOR III

Item number	Loading	Item
Age	0.63	Chronological age
7	0.52	I would (not) like a position which requires frequent changes from one task to another.
18	0.43	I (do not) like a great deal of variety in my work.
17	0.34	I dislike having to learn new ways of doing things.
39	0.33	I always put on and take off my clothes in the same order.
31	0.33	I (have not) taken a good many courses on the spur of the moment.
9	0.26	I find it easy to stick to a certain schedule once I have started on it.

sented. Like the long version, it is not unidimensional, but consists of three factors and certain items loading highly on more than one of these.

The original correlation between age and total score on the Wesley was 0.27, and between the Matrices and the Wesley, −0.29. It is now obvious that these correlations arise from responses to different items in the questionnaire. The present results agree fairly well with conventional stereotypes of rigid behavior: people of low intelligence acting in a rigid manner because they only see certain aspects of a situation, the elderly being unwilling to change long established patterns of behavior, and some people of all ages displaying an orderly detail-loving nature. Factor scores should reveal more information about each individual than did the total score which, though not unidimensional, has been shown previously to yield useful data. The factor scores provide, possibly for the first time, a means of measuring the extent to which different types of rigidity operate in an individual. It is not to be expected that a given piece of behavior will always stem from only one of these types of rigidity, but that in a given person at least these three factors will be combined in causing behavior, and that knowledge of the extent of the three types in any individual could be used, for example, in training procedures given to that person.

A factor analysis of the full battery of rigidity tests given to this sample will shortly be carried out, with these Wesley factor scores included in the matrix. In this way it should be possible to follow-up the three Wesley factors and link them to the different types of rigidity included in the battery. Later it is hoped to use the new Wesley subscores in industrial and experimental studies of learning.

SUMMARY

The Wesley Rigidity Inventory was completed by 200 subjects ranging in age from 20 to 81. The responses to 39 inventory items were intercorrelated and the resulting matrix factor analyzed by the centroid method. Age and scores on Raven's Progressive Matrices 1938 and Mill Hill Vocabulary (synonyms) were included in the matrix. Three orthogonal factors were found in the inventory, with age loading 0.63 on one factor and the Raven score −0.34 on another. It was found that the short form of the Wesley inventory devised by Zelen and Levitt also contains items from the three factors. It is proposed to use these new Wesley subscores in future work with the inventory.

REFERENCES

CHOWN, SHEILA M. Rigidity—a flexible concept. *Psychol. Bull.*, 1959, **56**, 195–223.

JONES, M. B. Aspects of the autonomous personality. *US Naval Sch. Aviat. Med. res. Rep.*, 1954, Proj. No. NM 001058.25.16.

MCNEMAR, Q. On the number of factors. *Psychometrika*, 1942, **7**, 9–18.

RAVEN, J. C. *Progressive Matrices (1938)*. London: Lewis, 1938.

RAVEN, J. C. *The Mill Hill Vocabulary Scale*. London: Lewis, 1943.

WESLEY, ELIZABETH. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–134.

ZELEN, S. L., & LEVITT, E. A group level of aspiration technique as a measure of personality rigidity. *Proc. Iowa Acad. Sci.*, 1953, **60**, 569–572.

ZELEN, S. L., & LEVITT, E. Notes on the Wesley Rigidity Scale: The development of a short form. *J. abnorm. soc. Psychol.*, 1954, **49**, 472–473.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 495-496

INTROVERSION-EXTRAVERSION AND KINESTHETIC AFTEREFFECTS

ALLAN RECHTSCHAFFEN[1]
University of Chicago Medical School

AND

LAWRENCE J. BOOKBINDER[2]
Northwestern University

EYSENCK (1955) hypothesized that extraverts develop greater neural satiation than introverts and offered evidence which he interpreted as supporting this hypothesis. He used dual criteria of psychiatric diagnosis and scores on the Rhathymia (R) scale (Guilford, 1942) to define the introversion-extraversion dimension and amount of kinesthetic aftereffect (KAE) to measure neural satiation. His results showed a tendency for the extraverts to develop greater KAE. A more recent study (Rechtschaffen, 1958) disclosed no significant differences in amount of visual aftereffect (VAE) developed by introverted and extraverted college students, where R score alone was used to define the personality dimension. One possible explanation for the negative result in the latter study derives from the particular sense modality used in measuring aftereffect, i.e., a relationship between aftereffect and introversion-extraversion might have been obtained had KAE been used rather than VAE. The present study was designed to explore this possibility by comparing the KAEs of introverted and extraverted college students.

METHOD

Subjects. The R scale was administered to a large introductory psychology class. Students with the lowest R scores were requested to serve as the introvert subjects, those with the highest R scores as the extravert subjects. Few students refused to serve; thus, the experimental sample represents the vast majority of those asked to participate. The mean R score of the extravert group ($N = 37$) was 53.9; the *SD* was 4.9, and the range 48 to 69. The mean R score of the introvert group ($N = 37$) was 26.7; the *SD* was 4.9, the range 16 to 34. There were 17 males and 20 females in the extravert group and 21 males and 16 females in the introvert group.

Apparatus. The apparatus used to measure KAE was similar to the one used by Eysenck (1955). It consisted of a comparison scale, a test object, and two stimulus objects; all were blocks of polished hardwood. The comparison scale was 30 in. long, 2 in. deep, and tapered from 4 in. to 0.5 in. The stimulus and test objects were each 6 in. long and 2 in. deep. One stimulus object was 2.5 in. wide and the other 0.5 in. wide, the test object was 1.5 in. wide.

Procedure. Subjects were tested individually and were blindfolded during the entire experiment. They were given five practice trials each, a trial consisting of four adjustments of the comparison scale, with a 1-minute rest between trials. The procedure for a practice trial was as follows: the subject stood in front of a table on which were mounted the comparison scale and the test object. The thumb and middle finger of the preferred hand were positioned in a movable rider on the comparison scale, so as to grasp the sides of the scale. The corresponding fingers of the other hand grasped the sides of the test object. Subjects were instructed to move the rider along the comparison scale until they judged that the width of the comparison scale was the same as the width of the test object. Four such adjustments were made in rapid succession. Starting position of the rider and distance of the comparison scale from the subject were systematically varied on successive adjustments to reduce position and movement cues.

After a 1-minute rest following the fifth practice trial, the subject was given the first preinspection trial, which was identical to a practice trial. Immediately following the preinspection trial, the comparison scale and test object were replaced by the 2.5-in. and 0.5-in. stimulus objects, respectively. Then for 40 seconds the subject vigorously rubbed the sides of the 0.5-in. stimulus object with the thumb and middle finger of his nonpreferred hand, and at the same time he rubbed the 2.5-in. stimulus object with the thumb and middle finger of his preferred hand. This inspection period was immediately followed by a postinspection trial, which also followed the same procedure as the practice trials. This procedure was intended to produce a "double satiation" effect: i.e., theoretically, stimulation of the test object hand by the smaller stimulus object should make the test object appear wider following satiation, and stimulation of the comparison scale hand by the wider stimulus object should make the comparison scale seem smaller than it did prior to inspection. Both changes in perception should work together to produce wider adjustments on the comparison scale following inspection.

Three more measures of KAE, each measure consisting of a preinspection trial, inspection, and a postinspection trial, were made during the experimental session. A 3-minute rest separated each of the four measures. Amount of KAE was measured as the absolute difference between average preinspection and postinspection adjustments of the comparison scale for each trial.

RESULTS

The results do not support the hypothesis that extraverts develop more KAE than introverts. Analysis of variance showed no significant difference between personality groups on amount of KAE over the four trials. Neither was there a significant difference between introverts and extraverts on practice or preinspection trials, thus eliminating the possibility that group differences in constant error confound comparisons of amount of KAE.

---

[1] A major portion of this research was done while the senior author was at the Chicago Branch of the Veterans Administration Neuropsychiatric Research Laboratory, Veterans Administration West Side Hospital.

[2] Now at Veterans Administration Hospital, Ann Arbor, Michigan.

For both personality groups, amount of KAE on the first trial was considerably larger than on the last three; the appropriate *F* ratio showed the decrease of KAE over successive trials to be significant at the .001 level. That this decrease results, at least in part, from incomplete dissipation of KAE during the 3-minute rest intervals is suggested by the failure of preinspection trials to return to original practice levels (*F* ratio for preinspection trials was significant at the .001 level). Thus, the results of the first KAE trial become crucial, since subsequent trials may become confounded with residues of undissipated KAE; large first trial KAEs may preclude large second trial KAEs because of increased preinspection adjustments, combined with relatively asymptotic postinspection adjustments. Therefore, introverts and extraverts were also compared with respect to first trial KAEs alone. The mean first trial KAEs for introverts and extraverts were .32 and .29 in. per adjustment, respectively. The *t* value for the difference was .66, a nonsignificant value.

Neither do the present data support Eysenck's hypothesis that extraverts develop more persistent aftereffects than introverts, since the interaction term for trials × personality group was not significant at the .05 level for either preinspection or postinspection trials.

A sex difference in KAE was found. Males had a mean first trial KAE of .36 in. as compared to .25 in. for females; the *t* value for the difference was 2.12, significant beyond the .05 level. In view of the above difference and the unequal number of males and females in the two personality groups, separate comparisons of introverts' and extraverts' KAEs were made for each sex. Mean KAE for male introverts was .40 in. as compared with .33 in. for male extraverts; the *t* value was .83, not significant. Mean KAE for female introverts was .27 in. as compared with .22 in. for female extraverts; the *t* value was .73, also not significant. Thus, although there were slightly different numbers of males and females in the two personality groups and a significant sex difference in KAE, these factors did not seem to alter the major finding that extraverts showed no evidence of more KAE than introverts.

SUMMARY

Eysenck's hypothesis that extraverts show greater figural aftereffects than introverts was investigated by comparing the kinesthetic aftereffects of 37 subjects with high scores on Guilford's Rhathymia scale (extraverts) with those of 37 subjects with low scores on the scale (introverts). The two personality groups did not differ significantly in amount of aftereffect. Several differences remain between the present study and Eysenck's earlier one, but the generality of the hypothesis that extraverts develop greater aftereffects than introverts is seriously questioned.

REFERENCES

EYSENCK, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.

GUILFORD, J. P. *An Inventory of Factors STDCR.* Beverly Hills, Calif.: Sheridan Supply, 1942.

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283–291.

(Received August 3, 1959)